

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 12.146 Seconds
(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773

Sequence: 1 GILVTVAPVAPVAPVSSGP.....KAGVTLHSVGPAAFGCTT 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	216	12.2	921	2 A48184	transcription init
2	216	12.2	921	2 A45183	TRP-associated fac
3	181.5	10.2	1275	2 T33369	hypothetical prote
4	176	9.9	2035	2 A40718	host cell factor C
5	173.5	9.8	528	2 I47141	gastric mucin (clo
6	169	9.5	3020	2 A43932	mucin 2 precursor,
7	167.5	9.4	549	2 C87719	protein R19.6 (im
8	166.5	9.4	825	2 T29634	hypothetical prote
9	165.5	9.3	2187	2 T30826	nascent polypeptid
10	164.5	9.3	752	4 A57784	AKM1/MTG8 mutant f
11	162	9.1	1367	1 S48478	glucan 1,4-alpha-g
12	158.5	8.9	1777	2 T34369	hypothetical prote
13	158	8.9	798	2 T34248	hypothetical prote
14	157	8.9	796	2 T21460	hypothetical prote
15	156	8.8	1324	2 S52863	DNA-binding protei
16	155.5	8.8	670	2 F36791	hypothetical prote
17	155.5	8.8	1151	2 T18535	high molecular mas
18	154.5	8.7	886	2 S29605	glycoprotein 350/2
19	154.5	8.7	3507	2 T34513	hypothetical prote
20	153.5	8.7	873	2 A47283	calphostin - fruit
21	153.5	8.7	2232	2 T34434	hypothetical prote
22	152	8.6	604	2 C57784	MTG8 protein splic
23	151	8.5	1169	2 S38181	flocculation prote
24	150.5	8.5	1630	2 A53577	ascites stialoglyco
25	150	8.5	865	2 A47282	calcium-binding pr
26	150	8.5	873	2 F96615	probable Myb-famil
27	146.5	8.3	5376	2 T42215	zonadhesin - mouse
28	146	8.2	495	2 A44489	GR box-binding pro
29	145.5	8.2	3570	2 T45025	mucin MUC5B, trach

30	143.5	8.1	294	2 A37232	mucin, tracheal (A
31	143.5	8.1	797	1 VGBEX1	glycoprotein X pre
32	143.5	8.1	1547	2 T28657	blackjack protein,
33	143	8.1	1161	2 S57180	probable membrane
34	142.5	8.0	780	2 A48143	HF-1 regulatory el
35	142.5	8.0	2476	2 T34022	zonadhesin - pig
36	142	8.0	725	2 A41258	aaglutinin core
37	141	8.0	660	2 JWM067	chitinase (EC 3.2.
38	141	8.0	867	2 T45463	membrane glycoprot
39	140	7.9	1216	2 F88473	protein F40H6.5 (I
40	139	7.8	851	2 T22696	hypothetical prote
41	138.5	7.8	770	2 T22808	hypothetical prote
42	138.5	7.8	1032	2 T34433	hypothetical prote
43	138.5	7.8	1832	2 T31113	mucin-like glycopr
44	138.5	7.8	2422	2 T12687	ALR protein homolo
45	137.5	7.8	1002	2 A56678	yemanuclein-alpha

ALIGNMENTS

RESULT 1

A48184 transcription initiation factor IID 110K chain - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 16-Feb-1994 #sequence.revision 18-Nov-1994 #text.change 15-Oct-1999

C:Accession: A48184

R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.

A:Title: The Drosophila 110-KDa transcription factor TFIID subunit directly interacts

A:Reference number: A48184; MUID:9317591; PMID:8327460

A:Accession: A48184

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-921 <KOK>

A:Cross-references: GB:563550; NID:9398432; PID:9398433

A:Experimental source: embryo nuclear extract

A:Note: sequence extracted from NCBI backbone (NCBI:134863, NCBI:134864)

C:Genetics:

A:Gene: FlyBase:Tafl10

A:Cross-references: FlyBase:FBgn0010280

C:Keywords: transcription initiation

Query Match	Score	216	DB 2	Length	921
Best Local Similarity	24.9%				
Pred. No. 4.7e-06					
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;					
QY 25 POIYAVKAPNTTIOFPANLQLPSTVLKNSGFLM-----VSPQOTVRAETTS 76					
DB 133 POSPEITLTNTGTGTPA-----LLVKTNDGFLRVGTGTPPVTOITWTSNNS 184					
QY 77 NITSBPANPAPQVYKICTVNSSQ-----LIKKVAVTPPKKLAQIGTVVTVTP 127					
DB 195 NTSSTNHPPTTQ-TRLOTPAPAAASMTNTATSNITVSVSSGVANSOPPHLTOLNAQ 243					
QY 128 KPSSVQSAVPVTVTPGKPLNTVTTLKPSLGSASTSPNEPLKKAENSAVAQINSP 187					
DB 244 APQLPQITQITPAGOSQOOQOVNVSAGTATVSTTA-----ATT 287					
QY 188 TMLENVK-KCKNFMLIKLACSGSQSPKMGONKVEQLLDKIEAEFTRLKLYELK 246					
DB 288 TQGGTKKCKRFLANLIEL--STREPKPVKKNVTLIOELVNVANVEDEECRDLRLIN 345					
QY 247 SSPQHVLPFLKSVVALROL-----LPSQSTIO--- 276					
DB 346 ASPQCLLGFLLKSLPLRLQALYKELVIEGIRKPPQVGLAGISQQLPIQAOIRIG 405					
QY 277 ---OCVOOTSSDWIATCTTVTTSPPVTTVSSSQSEKSIIVSGATAPRTVS---VQT 329					
DB 406 PSQTTTIGQTVRM--TPNALGTPRPTIGTTTISKQPN---IRLPAPALVNTGIRI 460					
QY 330 LNPPLAGPVGAKAGVTLHSVGPAA 354					

A:Reference number: A91831; MUID:87194600; PMID:3106330
 A:Accession: A26877
 A:Molecule type: DNA
 A:Residues: 1-242 <YAM>
 A:Cross-References: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
 A:Accession: B26877
 A:Molecule type: DNA
 A:Residues: 762-1331 <YA2>
 A:Cross-References: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
 R:Parlo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A>Title: Similar short elements in the 5' regions of the STRA and SGA genes from Sacchar
 A:Reference number: S27281; MUID:89031230; PMID:3141213
 A:Accession: S27281
 A:Molecule type: DNA
 A:Residues: 1-31 <PKR>
 A:Cross-References: EMBL:X13857; NID:g4551; PIDN:CAA2069.1; PID:g4552
 R:Lambrecht, M.G.; Bauer, F.F.; Marmur, J.; Precorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A>Title: Muc1, a mucin-like protein that is regulated by Mas10, is critical for pseudohy
 A:Reference number: JG6123; MUID:96323237; PMID:8710886
 A:Accession: JG6123
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1367 <LAM>
 A:Cross-References: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
 C:Genetics:
 A:Gene: SGD:WUC1; STA2; MAL5; DEX2; SGD:S0001458
 A:Cross-References: MIPS:YIR019C; SGD:S0001458
 A:Map position: 9R
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F:5-21/Domain: transmembrane #status predicted <TM1>
 F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 9.1%; Score 162; DB 1; Length 1367;
 Best Local Similarity 21.2%; Pred. No. 0.021;
 Matches 87; Conservative 62; Mismatches 181; Indels 80; Gaps 13;

```

QY 2 TLVTKVAPVSAPPKSSGPRLPAP-----QIVAAKAPNTTITQIPANLQIPGTVLIR 54
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 618 TTESSAPVTSTSTESSAPVPPTSSSTESSAPVPPTSSSTESSAPVPPTSSSTTE 677
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 55 SNGSPMLVSPOOT--VTRAETSNITSRPNANPQVTKICTVNSSQLIKKNAV-- 109
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 678 SSSAPVTSTSTESSAPVPPTSSSTESSAPVPPTSSSTESSAPVPPTSSSTESSAPV 737
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 110 -TPVKIAQIGTVVT-----VPRSSVQS-----VAVPTSVTVTPGKPLNT- 152
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 738 PTPSSSTESSAPVTSTSTESSAPVPPTSSSTESSAPVPPTSSSTESSAPVPPTP 797
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 153 -----VTTLKPSLGASTPNEPNLKAENSAVQINLSPTMLENVKCKNFIA 201
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 798 SSSSTESSAPVPPTSSSNITSSAPSTPSSSTESSAPV---PTPSSSTESS--- 851
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 202 MLIKLACSSQS-----PEMGQNKILVEQLDAKIAEETRLKLYELKSSPPHLY 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 852 --APVSSSTESSAPVPPTSSSNITSSAPSSIPSSSTESSTGTTPVSSK----- 904
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 255 PFLKRSVALROLPLPNSGFIOCVQOTSSDMVIACTTCTVTSP---VTTTVSSQSE 311
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 905 -----YRGSQR--ETSVSTETETIVPKITTSVTPSTTTTITTVCTGTIN 949
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 312 KSIIVSGATAPRTV--SVQTLNPLAGPVGAKAGVTLHSGVPTAANGTT 359
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 950 SAGETTSKSPKVTVTVTPTTTTSTTTTSTTTTTCSTGNSAGETT 999
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 12
 T34369
 hypothetical protein T19D12.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34369
 R:Favella, A.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid T19D12.
 A:Reference number: 221513
 A:Accession: T34369
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1777 <FAV>
 A:Cross-References: EMBL:U41263; PIDN:AAC24428.1; GSPDB:GN00020; CESP:T19D12.1
 A:Experimental source: strain Bristol N2; clone T19D12
 C:Genetics:
 A:Gene: CESP:T19D12.1
 A:Map position: 2
 A:Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1

Query Match 8.9%; Score 158.5; DB 2; Length 1777;
 Best Local Similarity 24.0%; Pred. No. 0.049;
 Matches 94; Conservative 51; Mismatches 158; Indels 89; Gaps 17;

```

QY 3 LTVKAPVSAPPKSSGPRLPAPQIVAAKAPNTTITQIPANLQIPGTVLIRKNS--GP 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 222 VTTTVAQTTPAIVTTAN--TTQGVTTTACGVTTVYRQNSFLAFTTAPSTNTTQGV 279
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 LMLVSPQOQVTRAE-----TTSNITSRPANPQVTKICT-----VPNS-----S 101
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 280 TTTVGKTTTITVTAQNSTMAATTTASNTTQPVVTTSTTQISTTTAAQATPSSSVIPTT 339
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 102 QLIKKAIVAPVKKLAQIGTVVTVTPKPSVO--SVAVP--TSVAVTTPGKPLNTVTTIK 157
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 340 QTTQRPSTICIPSTVSTQCTSTSPITQVSSNPSSTYSNFTPSPTTLLT-STIA 398
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 158 PSSIGA--STPNEPNLKAENSAVQINLSPTMLENVKCKNFIAMLIKLACSSQSPM 216
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 399 PSTQGVPTSSKSSPN-----STPTTTTTP-----GASSSTL 430
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 217 GQVKKLVEQLDAKIAEETRLKLYELKSSPPHLYPFL-----KSYVALROL 268
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 431 GSSSTSTVSTTTPS-----TPKVSSTLTSQSPPTSPPLVSSSSGSSSTVVTSTIT 484
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 269 PNSQSF--IOCVQOTSSDMVIACTTCTVTSPVTVTVSSSQSEKSLIYGATAPRTV 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 485 PSTQGVPTSSKSPSTPSTSPPTPKSTVNSP--STTGATSTASPTTSSAPTSQSHSP 542
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 328 QTLNPLAGPVGAKAGVTLHSGVPTAANGTT 359
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 543 SS-----TMTSTVPTSTFAST 559
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 13
 T34248
 hypothetical protein F31D5.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34248
 R:Milcox, L.
 submitted to the EMBL Data Library, June 1995
 A:Description: The sequence of C. elegans cosmid F31D5.
 A:Reference number: 221494
 A:Accession: T34248
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-798 <WLD>
 A:Cross-References: EMBL:U28941; PIDN:AAC71100.1; GSPDB:GN00020; CESP:F31D5.4
 A:Experimental source: strain Bristol N2; clone F31D5
 C:Genetics:
 A:Gene: CESP:F31D5.4
 A:Map position: 2
 A:Introns: 63/1; 95/2; 122/3; 189/2; 309/2; 350/3; 364/3; 411/2; 414/1; 443/1; 543/3;

Query Match 8.9%; Score 158; DB 2; Length 798;
 Best Local Similarity 23.8%; Pred. No. 0.019;
 Matches 79; Conservative 46; Mismatches 125; Indels 82; Gaps 14;

64 SPOOTYRAETTSNITSRPA---VPANQYKICTVPSNSSLIKKAVATPVYKLAQGT 120
DB 221 TPSTSEASSTVSTTARTTMTIPPTPTTTIA---STSTVSTVSTVSTVSTVPT 276
QY 121 TVVTVTPKPSVQSVAVPTSVVTPGKPLNTV-----TLKPSLGASSTPSNEPNL 173
DB 277 TVVTVTPKPSVQSVAVPTSVVTPGKPLNTV-----TLKPSLGASSTPSNEPNL 336
QY 174 KA-----ENSAVOINLSPTMLENVKCKNFLMLIKLACSGSQSPKMGONVYK 222
DB 337 TATVATPLTKNLSDNSMAILHND-NMLYGIKLSRNGROYSPRTDRYLENAKFGYQOK 395
QY 223 LVEOLLAK-----IEAEFTKRLVYELKSSQPHLVPLPKSVALLROLLENSQ 272
DB 396 I--RLKLSRRRMRWGTCGFIQPAETV-----PPPR-----VPRSA 428
QY 273 SFIOQCVQO---TSSDVIATCT--TWTVSPPVTVTTVSSQSEKSIIVSGATAPRVSV 327
DB 429 GFYAEFYQNMATISQIVPLSCAQTISLPPTVMSATVMMNGVAVSV---SATSEATMAL 485
QY 328 QTLN-----PLAGPVGAKAGVATLHSGVPTA 353
DB 486 LRFNISTLMKPLHGNL-----KYTWMDIGRNA 512

RESULT 14
T21460
hypothetical protein zk945.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21460
R:Wilkinson, J.
submitted to the EMBL Data Library, March 1995
A:Reference number: 219425
A:Accession: T21460
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-796 <NLE>
A:Cross-references: EMBL:Z48582; PIDN:CAA84469.1; GSPDB:GN00020; CESP:ZK945.10
A:Experimental source: clone F27E5
C:Genetics:
A:Gene: CESP:ZK945.10
A:Map position: 2
A:Insertions: 62/3; 124/3; 169/1; 216/1; 262/3; 371/1; 581/3; 608/1

Query Match
Best Local Similarity 8.9%; Score 157; DB 2; Length 796;
Matches 84; Conservative 64; Mismatches 188; Indels 74; Gaps 13;

QY 2 TLVTKVAPVSAAPKVVSGPRLPAPOIVAANKAPNTTIOFPANLQLPQTVLTKNSGFLM 61
DB 356 TLSTSIPTTTPTTSTSLSLSDNALCSYLDFTTSTFTTMTLTSTTEPSTSTTTE 415
QY 62 LVSTQCVTVTAETTSNITSRAVPA--NPQTVKICTVPSNSSLIKKAVATPVYKLAQIG 119
DB 416 VTSTSTVTPTEPTTTLTSTASTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 465
QY 120 TVVTVTPKPSVQSVAV-PTSVAVTVPKPLNTVTLKPSLGASSTPSNEPNL 178
DB 466 SSSSTVTTTPTST 521
QY 179 A-AVOINLSPTMLENVKCKNFLMLIKLACSGSQSPKMGONVYKLEOLLDA----- 230
DB 522 ASSTQSTSTQOOSTTTSKSE-----TTSSTQSTNDPF-YFVKATTTTFDSTSVNLT 573
QY 231 -----KIEAEFTKRLVY-----LKSSPQPHL-----VPLKSVALLR 265
DB 574 NSGLGIGYQTSIECTPSISNVSTTKDGACFTKSVSMPLGCTYPASTFVCGNYTFR 633
QY 266 QLLPNS-----OSFIQCVQOOTSMDVIANCTTTVTSP---VTTTVSSSQ 309
DB 634 ATMTTDDKKVYYTANYTIOEYSTTIESSTISAVASTSTSTSTSTSTSTSTSTST 693

QY 310 SEKSIIVSGATAPRTVSQVOTLNPLAGVYKAGVATLHSGVPTATGCTT 359
DB 694 STRSSDSSTTTAGSTTTTQESTTSTSEESTDSTSTTISTSTSTSSPST 743

RESULT 15
552863
DNA-binding protein R kappa B - human
C:Species: Homo sapiens (man)
C>Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 01-Dec-2000
C:Accession: S52863; A45580
R:Netters, A.; Boumeester, T.; Scheiderelt, C.
submitted to the EMBL Data Library, August 1994
A:Reference number: S52863
A:Accession: S52863
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1324 <NLE>
A:Cross-references: EMBL:X80878; NID:9695578; PID:9695579
R:Adams, B.S.; Leung, K.Y.; Hanley, E.W.; Nabel, G.J.
New Biol. 3, 1063-1073, 1991
A:Title: Cloning of R kappa B, a novel DNA-binding protein that recognizes the Interl
A:Reference number: A45580; MUID:92135142; PMID:1777480
A:Accession: A45580
A:Molecule type: mRNA
A:Residues: 313-1292, 'ISNREQLAP', 1304, 'QSPDLL', <ADA>
A:Cross-references: GB:008191; GB:S79520; NID:9476273; PIDN:AA17871.1; PID:9476274
A:Note: sequence extracted from NCBI backbone (NCBIIN:79520)
C:Keywords: DNA binding; transcription factor

Query Match
Best Local Similarity 8.8%; Score 156; DB 2; Length 1324;
Matches 99; Conservative 64; Mismatches 142; Indels 134; Gaps 21;

QY 4 VTKVAPVSAAPKVVSGPRLPAPOIVA---VKAPNTTIOFPANLQLPQTVL----- 52
DB 830 VRVVAQPSLPAVPOQSGPQOTLPMPAGPQIRPATAT-----QTKVVPQTVATVYK 884
QY 53 IKSNSGPMVLVSPQO---TWTVAETTSNITSRAVPA---PANQTVKICTVPSNSSLQ-LI 104
DB 885 AQTAAIVQRPQGTGLTSLPATASPVSKPATSSPGTSAPASTAAVIAQNTGONTI 944
QY 105 KKVAVT---PVKLAQIGTVVTVTPK-----PSSVQSVAVPTSVVTPGK- 148
DB 945 KOVAITGQLGVK--PQGNISLPLATNFRIGKQVLRPLPESITTTDAKGTQVLRITPDM 1002
QY 149 ---PLNVTYTK-----PSSLGASSTPSN--- 169
DB 1003 ATLAQSOVTVKTLQDLEGTGNTTGKISATLHVTSPVHAADSPAKASSASAPSTPT 1062
QY 170 -----EPNLKANSAAVOINLSPTMLENV--KCKNFL-----AMLIKLACS-GSQ 212
DB 1063 GTTVVATVPDLKPTPEASAPRLMPALGVSVADKGSSTVASSAKKAPATATRIYQGLGV 1122
QY 213 SPKMGONVYKLEOLLDAKIEAEFTKRLVYELKSSQPHLVPLPKSVALLROLLENSQ 272
DB 1123 PPKAGQRTTATVTHAKQGSVASSGSGT-----VHNSAVS-----LPSMN 1160
QY 273 SFIOQCVQO---QTSMDVIANCTTTVTSPVTVSSSQSEKSIIVSGATAPRTV--- 325
DB 1161 AAVSKTVAVASGAASTPISISTGAPTVKQVSVSTVSTVSTQAGK--LPTRITVPLSVISQ 1218
QY 326 -----SVQTLNPLAGPVCA 339
DB 1219 PMKGSVATADIKGNLGA 1237

Search completed: February 16, 2003, 22:00:49
Job time : 21.146 secs

Gencore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 6.91863 Seconds

(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773
Sequence: 1 GLVTVKAPVAPSVAPKVVSSGP.....KAGVTLHSVGPAAAGCTT 359

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1773	100.0	1 T2DT_HUMAN	092750 homo sapien
2	491	27.7	1 T2D3_HUMAN	000268 homo sapien
3	216	12.2	1 T2D3_DROME	P47823 drosophila
4	177.5	10.0	1 HFC1_MESAU	P51611 mesocricetu
5	175	9.9	1 HFC1_HUMAN	P51610 homo sapien
6	167.5	9.4	1 MUC2_HUMAN	002817 homo sapien
7	162	9.1	1 AMYH_YEAST	P08640 saccharomyc
8	157	8.9	1 K89_CABEL	Q09624 caenorhabdi
9	155.5	8.8	1 V630_HSV1	Q00130 ictaluriid h
10	154.5	8.7	1 V6P3_EBYA8	Q07284 epstein-bar
11	153	8.6	1 SP2_HUMAN	002086 homo sapien
12	152	8.5	1 MTG8_HUMAN	006455 homo sapien
13	151	8.5	1 YK82_YEAST	P36170 saccharomyc
14	150	8.5	1 CPN_DROME	Q02910 drosophila
15	147	8.3	1 ZAN_MOUSE	Q61909 mus musculu
16	146.5	8.3	1 ZAN_MOUSE	088799 mus musculu
17	145.5	8.2	1 PODX_HUMAN	000592 homo sapien
18	143.5	8.1	1 VGLX_HSVB	P28968 equine hebp
19	143	8.1	1 DAN4_YEAST	P47179 saccharomyc
20	142.5	8.0	1 ZAN_PIG	Q28983 sus scrofa
21	142	8.0	1 AGAL_YEAST	P32323 saccharomyc
22	140	7.9	1 YPX5_CABEL	Q09277 caenorhabdi
23	139.5	7.9	1 AL51_CANAL	P46590 candida alb
24	137.5	7.8	1 YEMA_DROME	P25992 drosophila
25	137	7.7	1 YJH8_YEAST	P47033 saccharomyc
26	136.5	7.7	1 CH12_COCIM	P54197 coccidioidi
27	135.5	7.6	1 PCT1_PIG	Q29076 sus scrofa
28	135.5	7.6	1 1858	P54674 dictyosteli
29	134.5	7.6	1 VGP3_EBY	P33200 epstein-bar
30	134	7.6	1 FLO5_YEAST	P38894 saccharomyc
31	133.5	7.5	1 MUC1_XENLA	005049 xenopus lae
32	132.5	7.4	1 YM96_YEAST	Q04893 saccharomyc
33	132	7.4	1 ALS3_CANAL	Q74623 candida alb

34	131	7.4	1036	1	P200_MYCPN	P75211 mycoplasma
35	128.5	7.2	743	1	OCT1_HUMAN	P14859 homo sapien
36	128	7.2	3866	1	HRX_MOUSE	P55200 mus musculu
37	127.5	7.2	1574	1	SYJ1_RAT	Q62910 rattus norv
38	127	7.2	634	1	HMP1_CANAL	P46593 candida alb
39	127	7.2	1306	1	MSB2_YEAST	P32334 saccharomyc
40	127	7.2	2700	1	ZAN_HUMAN	Q39493 homo sapien
41	126.5	7.1	784	1	SP4_HUMAN	Q02446 homo sapien
42	125.5	7.1	1794	1	YAV1_SCHPO	Q10172 schizosacch
43	125	7.1	1048	1	P100_HCMVA	P08318 human cytom
44	125	7.1	1322	1	YAG3_YEAST	P39712 saccharomyc
45	124.5	7.0	1147	1	SRE1_HUMAN	P36956 homo sapien

ALIGNMENTS

RESULT 1
ID T2DT_HUMAN STANDARD; PRT; 801 AA.
AC Q92750;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=97011146; PubMed=8858156;
RA Dikstein R., Zhou S., Tjian R.;
RT hTAFII130.";
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC EMBL: Y09321, CAA70499.1; -
CC Genbank: HGNC:11538; TAF4B.
CC MIM: 601689; -
CC DR InterPro: IPR003894; TAF_hom.
CC DR SMART: SM00549; TAFH; 1.
CC KW Transcription regulation; Nuclear protein.
CC FT NON TER
SQ SEQUENCE 801 AA: 85658 MW: D12B4932EFA9CD2 CRC64;
Query Match 100.0%; Score 1773; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 2, ze-100;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GLVTVKAPVAPSVAPKVVSSGPRLPAQIVAKAPNTTTTQFPANLQLPQGVLIKNSGPL 60
|||||

```

Db 1 GRLVTVAPVAPSAAPKSSSGRRLPAPQIVANVANKAPNTTITIOFPANLOLEPGTVILKNSGRL 60
QY 61 MLVSPQOQVTRAFETNTSNTSRPAPVNPQVTKICTVNPSSSOLIKKVAATPVKKLAQIGT 120
Db 61 MLVSPQOQVTRAFETNTSNTSRPAPVNPQVTKICTVNPSSSOLIKKVAATPVKKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAVPVTVTPGKPLNTVTTTLKPSLSGASSPSNPENKAKESNA 180
Db 121 TVVTVTPKSSVQSVAVPVTVTPGKPLNTVTTTLKPSLSGASSPSNPENKAKESNA 180
QY 181 VOINLSPTMLENVKCKNFAMLIKLAGSGSSQSPENQONKAKLVEQLLDKIAEETFRK 240
Db 181 VOINLSPTMLENVKCKNFAMLIKLAGSGSSQSPENQONKAKLVEQLLDKIAEETFRK 240
QY 241 LVEELKSSPQPHLVPLFLKSSVALROLPLNSQSFIOCCVOQSSDMVIATCTTTVTTSV 300
Db 241 LVEELKSSPQPHLVPLFLKSSVALROLPLNSQSFIOCCVOQSSDMVIATCTTTVTTSV 300
QY 301 VTTTVSSQSEKSIIVSGATPRTVSQVOTLNLAPGVGAKGVVTLHSVGPNTAATGTT 359
Db 301 VTTTVSSQSEKSIIVSGATPRTVSQVOTLNLAPGVGAKGVVTLHSVGPNTAATGTT 359

```

RESULT 2

T2D3_HUMAN STANDARD; PRT; 1083 AA.

```

ID T2D3_HUMAN 000268; Q99721; Q9BR40;
AC 000268; Q99721; Q9BR40;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor TFIIID 135 kDa subunit (TAFII-135)
DE (TAFII135) (TAFII130) (TAFII130).
GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97336072; PubMed-9192867;
RA Mengus G., May M., Carre L., Chamdon P., Davidson I.;
RT "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
RT mammalian cells."
RL Genes Dev. 11:1381-1395(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-21638749; PubMed-11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasalho M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurtry A.D.,
RA Milne S.A., Misty D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sulston J.E.,
RA Swann R.M., Symmons N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;

```

```

RT "The DNA sequence and comparative analysis of human chromosome 20."
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE-97098442; PubMed-8942982;
RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT "Molecular cloning and analysis of two subunits of the human TFIIID
RT complex: hTAFII130 and hTAFII100."
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
CC -1- FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
CC BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
CC -1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TAF6 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y11354; CAA72189.1; -
CC EMBL: AL137077; CAC36006.1; -
CC EMBL: AL109911; CAC22312.2; -
CC EMBL: U75308; AAC50901.1; -
CC TRANSFAC: T02328; -
CC Genew; HGNC:11537; TAF4.
CC MIM: 601796; -
CC InterPro: IPR003894; TAF_hom.
CC SMART: SM00549; TAFH; 1.
CC Transcription regulation, Nuclear protein.
KW DOMAIN 39 42
FT DOMAIN 52 57 POLY-ALA.
FT DOMAIN 98 101 POLY-GLY.
FT DOMAIN 142 148 POLY-ALA.
FT DOMAIN 268 275 POLY-PRO.
FT DOMAIN 331 337 POLY-ALA.
FT DOMAIN 680 683 POLY-PRO.
FT DOMAIN 808 813 POLY-ALA.
FT DOMAIN 828 831 POLY-ASP.
FT DOMAIN 828 831 POLY-ASP.
FT DOMAIN 105 117
FT CONFLICT 136 136 A -> S (IN REF. 2).
FT CONFLICT 185 185 G -> GPG (IN REF. 2).
FT CONFLICT 233 264 MISSING (IN REF. 3).
FT CONFLICT 293 293 P -> L (IN REF. 3).
SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;
Query Match 27.7%; Score 491; DB 1; Length 1083;
Best local Similarity 38.4%; Pred. No. 1,4e-22;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTITIOFPANLOLEPGTVILKNSGRL 61
Db 414 PRLTSSGIRATLTPVLAAPRLPQ-----PQNPNTIQ---NQQLPFGMVLVSENGQL 464
QY 62 LVSPQOQVTR-----AETTSNTSRPAPVNPQVTKICTVNPSSSOLIKKVAATPVKK 115
Db 465 MI-POALOMQAOAAHOPTMAAPAPPTSPAPVOISTVQAPGPIIAR-QVTP----- 518
QY 116 AQGTGVTVTPRPSSVQSVAVPTSVTPGKPLNTVTTTLKPSLSGASSPSNPENKAKES 168
Db 519 ----TIIKQV---SQAQTVQPSATLQSRPGVQPOLVLGAAGTASLGTAVVQGTGP 571
QY 169 NE-PNKAENSAVQINLSPTMLENVKCKNFAMLIKLAGSGSSQSPENQONKAKLVEQL 227
Db 572 RTVPGATTTSSAATE-----TMENVKCKNFSTILIKLAGSSQKOSTETRANVKEVONL 625

```


QY 228 LDATIEAEETFRKLYELKSSPOPHLYPLFKKSVVALROLIPNSQSFIOQCVOQ---TS 283
 DB 626 ILEEDTETSRILKSLNPPOLYPLFKSLPALROLPPDSAAFIQSOQOOPPTTS 685
 OY 284 SDWIACTTTTSPVTTTSSQSEKSIIVSGATAPRTVS 327
 DB 686 Q-----ATTALTAVALSSVORTACKTAFTVTSALQPVLSL 722
 RESULT 3
 T2D3 DROME STANDARD; PRT; 921 AA.
 AC P47825; P49845; Q9YU7;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription initiation factor TFIID 110 kDa subunit (P110)
 GN TAF110 OR CG5444.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID:7227;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 519-540; 597-616 AND 857-874.
 RC TISSUE-Embryo;
 RX MEDLINE-93145326; PubMed-7678780;
 RA Hoeij T., Weinzierl R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,
 RT "Molecular cloning and functional analysis of Drosophila TAF110
 RT reveal properties expected of coactivators.";
 RL Cell 72:247-260(1993).
 RT [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 398-406; 520-540 AND 860-877.
 RC TISSUE-Embryo;
 RX MEDLINE-93317591; PubMed-8327460;
 RA Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.,
 RT "The Drosophila 110-kDa transcription factor TFIID subunit directly
 RT interacts with the N-terminal region of the 230-kDa subunit.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).
 RT [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
 RA Botkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadiot E., Center A., Chandra I.,
 RA Cherry J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA DePamphilis M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson R., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harits N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei V., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleby J.M.,
 RA Palazzo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter G.M., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
 CC CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
 CC AND REPRESSORS. MAY FUNCTION AS A COACTIVATOR BY SERVING AS A SITE
 CC OF PROTEIN-PROTEIN CONTACT BETWEEN ACTIVATORS LIKE SPI (OR BTB)
 CC AND TFIID COMPLEX.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 DR EMBL; L06861; NOT_ANNOTATED_CDS.
 DR EMBL; S63550; AAB27433.1; .
 DR EMBL; AE003528; AAF49536.1; .
 DR TRANSFAC; T02121; .
 DR FlyBase; FBgn0010280; TAF110.
 DR InterPro; IPR003894; TAF_hom.
 DR SMART; SM00549; TAFH; 1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 66 82 POLY-GLN.
 FT DOMAIN 108 111 POLY-GLN.
 FT DOMAIN 259 265 POLY-GLN.
 SO SEQUENCE 921 AA; 99338 MW; 27E6852659872767 CRC64;
 Query Match 12.28; Score 216; DB 1; Length 921;
 Best Local Similarity 24.9%; Pred. No. 5,4e-06;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;
 OY 25 PQIYAVKAPNTTTOFPPNQLPPGTVLKNSGPIML-----VSQQTVAERTS 76
 DB 133 PQSPTIILSTINTGOTPA-----LIVKTDNGQLRVGTGGPPVTOTITNTSNN 184
 OY 77 NITSRPAVPANPQVVKICTVFNSSSQ-----LIKKVAVTPVKILQIGTVVTVTP 127
 DB 185 NTSSTNHPPTTQ-IRQTVPAAASMTNTATSNIVNVAASGVANSQPHPLQLQNAQ 243
 OY 128 KPSSVQSAVPTSVVYTPGKPLNTVTTLKPSISLGASSTPSNEPNLAENSAVAQINISP 187
 DB 244 APOLPQITQIDITIPAOQSOQOQVNNVSSAGCTATAVSTTA-----ATT 287
 OY 188 TMLENVK-KCKNPLAMIKLACSGSOSPENGONKVLVEOLLDAKIEEFTFRKLYELK 246
 DB 288 TQGTNKKCKKFLANLEL--STREKPKVKNVTTLIQELVANVEDEECRDERLLN 345
 OY 247 SSPPHLPLFLKKSVALROL-----LPSQSFTQ--- 276
 DB 346 ASPPCILGFLKSLPLRQLALYRKELVIEIKPPQHVGLAGLSQPLRIQAOIRIG 405
 OY 277 -----QCVOQISSDWIATCTTTTSPVTTTSSQSEKSIIVSGATAPRTVS---VOT 329
 DB 406 PSQTTTIGQITQVIRMI--TPNALGTPRPRTIGTTTISKOPPN---IRLPAPRLVTVGTGIRT 460
 OY 330 LNPAGPVGAKAGVTLHSVGPTAA 354
 DB 461 QIP-SLOYPGQANIVQIR--GPQHA 482

```

RESULT 4
HFC1_MESAU STANDARD: PRT; 2090 AA.
ID HFC1_MESAU
AC P51611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF)
(GFP).
GN HFC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
NC NCB1_TaxId=10036;
RN 111
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSLACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PPCE/THET, WITHIN THE HCF REPEAT (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC CC EMBL: D45419; BAA08258.1;
CC DR InterPro: IPR003961; FN_III.
CC DR InterPro: IPR001798; Kelch.
CC DR Pfam: PF00041; fn3; 2.
CC DR Pfam: PF01344; Kelch; 5.
CC DR SMART: SM00060; FN3; 2.
CC DR Nuclear protein; Repeat.
CC FT REPEAT 44 89
CC FT REPEAT 93 140
CC FT REPEAT 148 194
CC FT REPEAT 217 265
CC FT REPEAT 266 313
CC FT REPEAT 313 366
CC FT DOMAIN 1010 1448
CC FT REPEAT 1010 1035
CC FT REPEAT 1072 1097
CC FT REPEAT 1101 1126
CC FT REPEAT 1157 1182
CC FT REPEAT 1295 1320
CC FT REPEAT 1323 1348
CC FT REPEAT 1358 1383
CC FT REPEAT 1423 1448
CC FT REPEAT 1423 1448
CC SEQUENCE 2090 AA; 214942 MW; E495EB8B1F285E17 CRC64;

Query Match 10.0%; Score 177.5; DB 1; Length 2090;
Best Local Similarity 23.6%; Pred. No. 0.0031;
Matches 106; Conservative 60; Mismatches 144; Indels 139; Gaps 22;

```

```

QY 10 VSAPPKVSGPRLPAPQIVAVKAPNTTIOFPANLQLPSTVLIKNSGPMLVSPQQT 69
D 537 IGSNPMQSGMAALAAAAATOKITPSSA---PVLVSVPACTTIVKT-----VAVTPQT 588
QY 70 TRATTSNITSRAPVAPNPOT-----VKICVPRSSSO---LKKVAVTPYKRL 115
D 589 LPA--TVKVASPVMWSNPRTRMLKTAQAQVGSVSSAANTSTRPITTVKSGTVYTAQ 646
QY 116 AQIGTVV-----TVVPRP-----SSQSVAVPTS-----140
D 647 AQVTVTVGSGVTITLVKSPISVPGSALISNLGKMSVVOCKRPQVTSQAQSGPV 706
QY 141 -----VTVTPGKPLNVTTL-----KPSLSGASST-----PSNE 170
D 707 TQIIQTKGRLPAGTILKLVTSADGKPTTITVQAQSGAGSKPILIGSVSPSTRKPGT 766
QY 171 PNKAENSAVQNLSTPTLENVKKCNFLAML-IRKACGSGSPENGVAKRLVQOLD 229
D 767 TITKTPMSALITQAGATGVTSPGIKSPITITTKVMISGTAP-----AKITAV-- 818
QY 230 AKIEAEETFKRL-YVELKSP-OP---HLVPLKKSVALROLPLNSQSFIOCVQOTS 283
D 819 PKATGCGQGVQVYLVKAGPQALIRVP-----MSGVRLVTVYSVAVKPAV----- 869
QY 284 SDNVIACTTTVTTSPTVTTTSSQSEKSIYSATAPRTVSQVLNPLAGVGAAGV 343
D 870 -TVLVKGTGVTVTLGTVGTSTST-----LAGAGHSTSA-----SLATPITLGT 916
QY 344 VTLHS--VGPTA-----ATGGT 359
D 917 ATLSQVINPALTIVSAQDTTLTAAGLT 945

RESULT 5
HFC1_HUMAN STANDARD: PRT; 2035 AA.
ID HFC1_HUMAN
AC P51610;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF)
(GFP).
GN HFC1 OR HCF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NC NCB1_TaxId=9606;
RN 111
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND PARTIAL SEQUENCE.
RA MEDLINE=93327419; PubMed=8392914;
RA Wilson A.C., Lamarco K., Peterson M.G., Herr W.;
RT "The VP16 accessory protein HCF is a family of polypeptides processed
RT from a large precursor protein.";
RL Cell 74:115-125(1993).
RN 121
RP SEQUENCE OF 65-2035 FROM N.A. (LONG AND SHORT ISOFORMS).
RC TISSUE=Petal brain;
RA MEDLINE=95130107; PubMed=7829097;
RA Fratini A., Faranda S., Redolfi E., Zucchi I., Villa A.,
RA Patrosio M.C., Strina D., Susani L., Vezzoni P.;
RT "Genomic organization of the human VP16 accessory protein, a
RT housekeeping gene (HCF1) mapping to Xq28.";
RL Genomics 23:30-35(1994).
RN 131
RP PROCESSING.
RA MEDLINE=96033796; PubMed=7590226;
RA Wilson A.C., Peterson M.G., Herr W.;
RT "The HCF repeat is an unusual proteolytic cleavage signal.";
RL Genes Dev. 9:2445-2458(1995).
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSLACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING

```


DT 01-FEB-1995 (Rel. 31, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
 DE glucosidase) (1,4-alpha-D-glucan glucosylase).
 GN STAL OR STA2 OR MAL5 OR YIR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 ON NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
 RX MEDLINE=87194600; PubMed=1106330;
 RA Yamashita I., Nakamura M., Fukui S.;
 RT "Gene fusion is a possible mechanism underlying the evolution of
 RT STAL.";
 RL J. Bacteriol. 169:2142-2149(1987).
 RN [3]
 RP SEQUENCE OF 1-31 FROM N.A.
 RC STRAIN=SPX101-1C;
 RX MEDLINE=89031230; PubMed=1141213;
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STA2 and SGA genes
 RT from Saccharomyces cerevisiae.";
 RL FEBS Lett. 239:179-184(1988).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -1- SIMILARITY: TO S.POMBE SPBC215.13.
 CC -1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z38061; CA86176.1; -
 DR EMBL: M16164; AAA35014.1; -
 DR EMBL: M16165; AAA35015.1; -
 DR EMBL: X13857; CAA32069.1; -
 DR PIR: B26877; B26877.
 DR PIR: A26877; A26877.
 DR PIR: S48478; S48478.
 DR SCD: S0001458; MUC1.
 KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KM Signal; Multigene family.
 FT CHAIN 1 21 POTENTIAL.
 FT DOMAIN 22 1367 GLUCOMYLAASE S1/S2.
 FT CARBOHYD 817 817 SER/THR-RICH.
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1367 AA; 136110 MW; 91C0E82BDB61AA9D CRC64;
 Query Match 9.18; Score 162; DB 1; Length 1367;
 Best Local Similarity 21.28; Pred. No. 0.016;
 Matches 87; Conservative 62; Mismatches 181; Indels 80; Gaps 13;
 Oy 2 TLVTKVAPVSPKPVSSGCPRLPAP-----QIVAVKAPNTTTOFANQLDPTGVLIK 54
 Db 618 TTSSSAPVSTSTSSSAPVPTPTSSSTSSSAPVPTPTSSSTSSSAPVPTPTSSSTTE 677

Oy 55 SNSGFLMLVSPQOT--VTRAETTSNITSRPVAMPANPQTVICVTPNNSOLIKRVAV-- 109
 Db 678 SSAPVSTSTSSSAPVPTPTSSSTSSSAPVPTPTSSSTSSSAPVPTPTSSSTSSSAPV 737
 Oy 110 -TPVKKLAQIGTVVTT-----VPKPSVQS-----YAVFTSVYVTPGKPLMT- 152
 Db 738 PTPSSSTSSSAPVPTPTSSSTSSSAPVPTPTSSSTSSSAPVPTPTSSSTSSSAPVPTP 797
 Oy 153 -----VTLKPSILGASPTSPNEPNLKAENSAVOINLSPTMLENVKCKKNFLA 201
 Db 798 SSSTSSSAPVPTPTSSSNTTSSAPSTPTSSSTSSSVV---PTSSSTSSS--- 851
 Oy 202 MLKILACSGSOS-----PENGQVKKLVQLDADKIAEAEFRKLYELKSSPQPLV 254
 Db 852 --APVSSSTSSSAPVPTPTSSSNTTSSAPSSIPFSSSTESFSGTVTPSSSK----- 904
 Oy 255 PFKKSVVALQOLPNQSFQOCQOQSSDMVATCTTWTGSP---VTTTVSSQSE 311
 Db 905 -----YFGSQT--ETVSSTETTVPTKTTTSVTPSTTTTITTVCSGTGN 949
 Oy 312 KSIIVSGATAPRTV--SVQTLNPLAGPVGAKAGVTLHSGPTAATGCTT 359
 Db 950 SAGETSCSKRTVTTTPTTTTTSVTTSSTTTTITTVCSGTGNSAGETT 999

RESULT 8
 ID YS89 CAEEL STANDARD; PRT: 3178 AA.
 AC Q09624; Q09625; Q0969D;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 GN Hypothetical protein ZK945.9 in chromosome II.
 GN ZK945.9/ZK945.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Wilkinson-Sproat J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Durbin R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z48544; CAB70192.1; JOINED.
 DR EMBL: Z48582; CAB70192.1; JOINED.
 DR EMBL: Z48582; CAB70201.1; -
 DR EMBL: Z48544; CAB70201.1; JOINED.
 DR WormPep: ZK945.9; CE25697.
 DR InterPro: IPR002111; Cat_channel_Tyrl.
 DR InterPro: IPR001024; Lipoygenase_LH2.
 DR InterPro: IPR000636; M-channel_nlg.
 DR InterPro: IPR000203; PKD_cys_rich.
 DR Pfam: PF00520; Ion_trans_1.
 DR Pfam: PF01477; PLAT_1.
 DR Pfam: PF01825; GPS_1.
 DR SMART: SMO0303; GPS_1.
 DR SMART: SMO0308; LH2_1.
 KW Hypothetical protein; Transmembrane.

FT DOMAIN 266 1196 SER/THR-RICH.
FT DOMAIN 1105 1241 GLY/SER-RICH.
FT DOMAIN 2071 2120 GRS.
FT TRANSMEM 51 30 POTENTIAL.
FT TRANSMEM 2139 2161 POTENTIAL.
FT TRANSMEM 2348 2367 POTENTIAL.
FT TRANSMEM 2390 2412 POTENTIAL.
FT TRANSMEM 2451 2468 POTENTIAL.
FT TRANSMEM 2483 2505 POTENTIAL.
FT TRANSMEM 2567 2589 POTENTIAL.
FT TRANSMEM 2836 2858 POTENTIAL.
FT TRANSMEM 2939 2961 POTENTIAL.
FT TRANSMEM 2976 2998 POTENTIAL.
FT TRANSMEM 3038 3060 POTENTIAL.
SQ SEQUENCE 3178 AA; 344726 MW; 8823943603666CD CRC64;

Query Match 8.9%; Score 157; DB 1; Length 3178;
Best Local Similarity 20.5%; Pred. No. 0.091;

Matches 84; Conservative 64; Mismatches 188; Indels 74; Gaps 13;

OY 2 TLVTKAPVAPKVPSSGRLPAPQIVAKAPNTTIOFANQLPFGVLIKNSGPLM 61
DB 356 TLSTSTPTTPTTPTTSTLSSLPNMAICSYLDETTSTTTMLTSTTEPSTSTTE 415
OY 62 LVSPQOVTAEETSNITSRPAVPA--NPQVKTCTVPNSSOLIKKAVTPVKLAQIG 119
DB 416 VITSTSTVTTEPTTLTSTASTSTTEPSTSTSTSTSTSTSTSTSTSTSTSTST 465
OY 120 TVVTVTPKPSVQAVAV-PTSVVTVTPKPLTVTLKPSLGASTPSENPNAENS 178
DB 466 SSSSTVTPTST 521
OY 179 A-AVOINLSPMLNKKKCNFLAMLIKACSGSQSPENQNKKLVQLLDA----- 230
DB 522 ASSTGSTSTQOOSTTKSET-----TTSDGTNDF-YFEKATTTFTYDSTVNLTL 573
OY 231 -----KIAEEFTKLYE-----LKSSPQPHL-----VFLKKSVALR 265
DB 574 NSGLGIIGVOTSIETCSPTSSNVSTTKQACFTKSVSMRLGCTYPASFVPGCNTFR 633
OY 266 QLLPNS-----QSFIOQCVOQTSMDVIATCTTVTTSPT--VTTTVSSSQ 309
DB 634 ATTTTDDKKVY 693
OY 310 SEKSIIIVGATAPRTVSVQTLNPLAGPVGAKAIVLHSGPTAAAGCTT 359
DB 694 STRSDSTSTTSAGSTTTLOESTTSEESTTDSSTTTTISDTSTSSPSSST 743

RESULT 9

VG50_HSV11 STANDARD; PRT; 670 AA.
AC 000130;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical gene 50 protein.
GN 50
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Ictalurid Herpes-like viruses.
OX NCBI_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Auburn 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M75136; AAA88153.1; -

DR PIR: F36791; F36791.

KW Hypothetical protein; Repeat.

FT REPEAT 143 158
FT REPEAT 171 186
FT REPEAT 200 214
FT REPEAT 215 233
FT REPEAT 234 252
FT REPEAT 253 268
FT REPEAT 279 293
FT REPEAT 294 309
FT REPEAT 320 334
FT REPEAT 335 349
FT REPEAT 362 376
FT REPEAT 377 391
FT REPEAT 392 406
FT REPEAT 407 421
FT REPEAT 422 436
FT REPEAT 437 452
FT REPEAT 464 477
FT REPEAT 478 493
FT REPEAT 504 517
FT REPEAT 518 531
FT REPEAT 532 545
FT REPEAT 546 559
FT REPEAT 560 573
FT REPEAT 574 587
FT REPEAT 588 601
FT REPEAT 602 615
FT REPEAT 616 629
SQ SEQUENCE 670 AA; 64174 MW; 2B64A781C519E8B4 CRC64;

Query Match 8.8%; Score 155.5; DB 1; Length 670;
Best Local Similarity 24.5%; Pred. No. 0.017;

Matches 91; Conservative 33; Mismatches 129; Indels 119; Gaps 17;

OY 16 VSSGRLPAPQIVAKAPNTT-----IOPANQLPFGVLIKNSGPLMVSPQOVT 70
DB 248 VTTTPAMPAG---ANDANTTTATPTGANDTANTVMPAGATDVVTTTPAM-----PT 297
OY 71 RAETSNITSRPAVPANPOTVKTCTVPNSSOLIKKAVTP-----VKLAQIG 119
DB 298 GANDTANIT--ATPAGANDTANTVMPAGATDV--VTTTPAMPAGANDTANTKPAGST 353
OY 120 TVVTVTPKPSVQAVAVP-----SVYTVPGKP---LNTVTLKRS-SLGASTPSNEP 171
DB 354 DVVTTTP-----AMPGATDVVTTTPAMPGTGATDVVTTTPAMPGTGATDVVTTT 405
OY 172 NLKAENSAVOINLSPMLNKKCNFLAMLIKACSGSQSPENQNKKLVQLLDAK 231
DB 406 PAKPAGANGVTTTPAM-----PAGAN----- 428
OY 232 IEAEFTKLYELKSSQPHLVFLKKSVALROLPLNSQSFIOQCVOQT-SSDMVIAT 290
DB 429 -----DIVVTTAPATPAGANDTANTVTKPTGATDVVTT 461
OY 291 CTT--TVTSPVYTVTVSSSQSEKSIIVSGAT-APRTVSVQTLNPLAGPVGAGAVYTLH 347
DB 462 ATVAKPTGATGTVTTTAKPTGANDTANTVTKPTGATGTVTTT---AKPTGA-TGTVYVA 516
OY 348 SVGPATAGGTT 359
DB 517 TAKPTGATGTVT 528

RESULT 10
VG03_EBVAB

or send an email to license@slb.ch.

CC EMBL: D28588; BA05923.1; -
 CC EMBL: M97190; AAA5629.1; -
 DR PIR: A44489; A44489.
 DR HSSP: P08047; 1SP2.
 DR TRANSFAC: T02356; -
 DR Genew: HGNC:11207; SP2.
 DR MIM: 601801; -
 DR InterPro: IP000822; Znf_C2H2.
 DR Pfam: PF00096; Zf-C2H2; 3.
 DR PRINTS: PR00048; ZINCINGER.
 DR ProDom: PD000003; Znf_C2H2; 2.
 DR SMART: SM00355; Znf_C2H2; 3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
 DR Transcription regulation: Activator; Zinc-finger; Metal-binding;
 KW DNA-binding; Nuclear Protein; Repeat.
 FT DOMAIN 518 600 ZINC_FINGERS.
 FT ZN_FING 518 542 C2H2-TYPE.
 FT ZN_FING 548 572 C2H2-TYPE.
 FT ZN_FING 578 600 C2H2-TYPE.
 SQ SEQUENCE 606 AA; 64153 MW; A27C6D46D36E186 CRC64;

Query Match Best Local Similarity 24.9%; Pred. No. 0.021; Score 153; DB 1; Length 606;
 Matches 96; Conservative 52; Mismatches 137; Indels 100; Gaps 19;

OY 7 VAVPVSAPVSSGPRLP-----DQIVAVKAPNTTIOFPANL-----QLPQGVLIKS 55
 DB 104 VFAIQPTMINKGTRSNANIOYQVPI---QASNOTIOVQNLNQIIRGT----- 155
 OY 56 NSGPIMLVSPQCTVTAETTSNTTSRPANPQVTKICVPPSSSOLLKKAIVTPYK-- 113
 DB 156 -----NOAITTPSPSHKPPVPIKPAPIO-----KST-----TTPVQSG 190
 OY 114 ----KLAQIGTVVTVTPKPSVOS--VAVPTSVTVTPGKPLN--TVTLKPSLSGSS 166
 DB 191 ANVVKILTGCGGANTLPLPANNVNASDTCAPQIOLTESPTPLSKNKAKRKKSLPASOP 250
 OY 167 PSNEPLKAENSAVOINLSPWLENVKKCKNEFLMLKLACSGSSQSPENQVKKLVEQ 226
 DB 251 P-----VAVAEQVETVLETTADNIIQGNLLIV-----QSPGGGQPAVVQVQ 295
 OY 227 LLDAKTEAEFFRKLYVELKSSPQPHLVPLKSSVALNQLPN-----SQSTIQCVQO 261
 DB 296 VVPPKAEQOQVVO-----IPQAL-----RVQAASTLPTVPQKPSQNPQIOAAEP 342
 OY 282 TSSDMYIAT-----CTTVTTSFVVT-----TVSSSSQSEKSIIVGATAPRTVSV--QTL 330
 DB 343 TPTQVYIRTPSGGVQVIVQDSPRAAATNSNTTCSPPASRAPHLSGTSKKHSAILRKE 402
 OY 331 NPLACPVGAKAG-VVTLHSGVPTAA 354
 DB 403 RPL--PKIAPAGSIISLNAQLAA 425

RESULT 12
 MTG8_HUMAN STANDARD; PRT; 604 AA.
 AC Q06455; Q06455; Q92479; Q16346; Q16347; Q14873; Q9BR20;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE MTG8 protein (ETO protein) (Eight twenty one protein) (Cyclin D
 DE related protein).
 GN CBFA2T1 OR MTG8 OR ETO OR CDR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT AML1-MTG8/ETO FUSION IN AML-M2.

RC TISSUE=Retal brain;
 RX MEDLINE=93327761; PubMed=8334990;
 RA Miyoshi H., Kozu T., Shimizu K., Enomoto K., Maseki N., Kaneko Y.,
 RA Kamada N., Ohki M.;
 RT "The t(8;21) translocation in acute myeloid leukemia results in
 RT production of an AML1-MTG8 fusion transcript.";
 RL EMBO J. 12:2715-2721(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98326315; PubMed=9661669;
 RA Wolford J.K., Prochazka M.;
 RT "Structure and expression of the human MTG8/ETO gene.";
 RL Gene 212:103-109(1998).
 RN [3]
 RP SEQUENCE OF 31-604 FROM N.A.
 RX MEDLINE=94185017; PubMed=8137293;
 RA Erickson P.F., Robinson M., Owens G., Drabkin H.A.;
 RT "The ETO portion of acute myeloid leukemia t(8;21) fusion transcript
 RT encodes a highly evolutionarily conserved putative transcription
 RT factor.";
 RL Cancer Res. 54:1782-1786(1994).
 RN [4]
 RP SEQUENCE OF 30-604 FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 334-432 FROM N.A., AND VARIANT AML1-MTG8/ETO FUSION.
 RX MEDLINE=95002916; PubMed=7919324;
 RA Tyhge J.E., Calabi F.;
 RT "Alternative, out-of-frame runt/MTG8 transcripts are encoded by the
 RT derivative (8) chromosome in the t(8;21) of acute myeloid leukemia
 RT M2.";
 RL Blood 84:2115-2121(1994).
 RN [6]
 RP VARIANT AML1-MTG8/ETO FUSION IN AML-M2.
 RX MEDLINE=95329434; PubMed=7541640;
 RA Era T., Asou N., Kunitada T., Yamasaki H., Asou H., Kamada N.,
 RA Nishikawa S.-I., Yamaguchi K., Takatsuki K.;
 RT "Identification of two transcripts of AML1/ETO-fused gene in t(8;21)
 RT leukemic cells and expression of wild-type ETO gene in hematopoietic
 RT cells.";
 RL Genes Chromosomes Cancer 13:25-33(1995).
 RN [7]
 RP VARIANT AML1-MTG8/ETO FUSION IN AML-M2.
 RX MEDLINE=93357484; PubMed=8353289;
 RA Kozu T., Miyoshi H., Shimizu K., Maseki N., Kaneko Y., Asou H.,
 RA Kamada N., Ohki M.;
 RT "junctions of the AML1/MTG8(ETO) fusion are constant in t(8;21) acute
 RT myeloid leukemia detected by reverse transcription polymerase chain
 RT reaction.";
 RL Blood 82:1270-1276(1993).
 RN [8]
 RP VARIANT AML1-MTG8/ETO FUSION IN AML-M2.
 RX MEDLINE=93046086; PubMed=1423235;
 RA Nilsson P.E., Watkins P.C., Sacchi N.;
 RT "Transcriptionally active chimeric gene derived from the fusion of
 RT the AML1 gene and a novel gene on chromosome 8 in t(8;21) leukemic
 RT cells.";
 RL Cancer Genet. Cytogenet. 63:81-88(1992).
 CC -1- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: MTG8A AND MTG8B (SHOWN HERE);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANTLY EXPRESSED IN BRAIN. LOWER
 CC LEVELS IN LUNG, HEART, TESTIS AND OVARY.
 CC -1- DISEASE INVOLVED IN ACUTE MYELOID LEUKEMIA (AML-M2) BY A
 CC CHROMOSOMAL TRANSLOCATION t(8;21)(Q26;Q22) THAT INVOLVES CBFA2T1
 CC AND PROTEIN AML1.
 CC -1- SIMILARITY: CONTAINS 1 MYND-TYPE ZINC FINGER.
 CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chromancer/Genes/ETO.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D14821; BAA03558.1; ALT_SEQ.
DR EMBL: D13979; BAA03089.1; ALT_SEQ.
DR EMBL: D14289; BAA03247.1; ALT_INIT.
DR EMBL: AF018282; AAC28932.1; JOINED.
DR EMBL: AF018271; AAC28932.1; JOINED.
DR EMBL: AF018272; AAC28932.1; JOINED.
DR EMBL: AF018273; AAC28932.1; JOINED.
DR EMBL: AF018274; AAC28932.1; JOINED.
DR EMBL: AF018275; AAC28932.1; JOINED.
DR EMBL: AF018276; AAC28932.1; JOINED.
DR EMBL: AF018277; AAC28932.1; JOINED.
DR EMBL: AF018278; AAC28932.1; JOINED.
DR EMBL: AF018279; AAC28932.1; JOINED.
DR EMBL: AF018281; AAC28932.1; JOINED.
DR EMBL: AF018282; AAC28931.1; JOINED.
DR EMBL: AF018270; AAC28931.1; JOINED.
DR EMBL: AF018272; AAC28931.1; JOINED.
DR EMBL: AF018273; AAC28931.1; JOINED.
DR EMBL: AF018274; AAC28931.1; JOINED.
DR EMBL: AF018275; AAC28931.1; JOINED.
DR EMBL: AF018276; AAC28931.1; JOINED.
DR EMBL: AF018277; AAC28931.1; JOINED.
DR EMBL: AF018278; AAC28931.1; JOINED.
DR EMBL: AF018279; AAC28931.1; JOINED.
DR EMBL: AF018281; AAC28931.1; JOINED.
DR EMBL: D43638; BAA07755.1; JOINED.
DR EMBL: BC005850; AAH05850.1; ALT_INIT.
DR EMBL: X79990; CA56311.1; JOINED.
DR EMBL: S74096; AAB32126.1; JOINED.
DR EMBL: S74092; NOT_ANNOTATED_CDS.
DR EMBL: S74158; AAB34819.2; ALT_SEQ.
DR EMBL: S78159; AAB34820.2; ALT_SEQ.
DR EMBL: D14822; BAA03559.1; ALT_SEQ.
DR EMBL: D14823; BAA03560.1; ALT_SEQ.
DR EMBL: S50186; NOT_ANNOTATED_CDS.
DR Genew: HGNC:1535; CBFA2T1.
DR MIM: 133435; JOINED.
DR InterPro: IPR003894; TAF_hom.
DR InterPro: IPR002893; znf_MYND.
DR Pfam: PF01753; zfp_MYND; 1.
DR SMART: SM00549; TAFH; 1.
KW Zinc-finger; DNA-binding; Nuclear protein; Alternative splicing;
KW Chromosomal translocation; Proto-oncogene; Transcription regulation.
FT DOMAIN 57 60 POLY-PRO.
FT DOMAIN 102 108 POLY-SER.
FT DOMAIN 290 297 POLY-PRO.
FT DOMAIN 408 413 POLY-SER.
FT SITE 30 31 BREAKPOINT FOR TRANSLLOCATION TO FORM
FT 2N-FING 515 551 AML1-MTG8 IN AML-M2.
FT VARSPLIC 1 29 MYND-TYPE.
FT SEQUENCE 604 AA; 67566 MW; C3D2452P96E5679 CRC64; MISVRNTWRLSLVIDCKRKKNFEYEQ -> MP (IN
SQ

Query Match 8.6%; Score 152; DB 1; Length 604;
Best Local Similarity 29.7%; Pred. No. 0.024;
Matches 66; Conservative 39; Mismatches 89; Indels 28; Gaps 10;

OY 89 QTVKICVPSNSOLIKKVAATPVKAKLAQIGTVVTVPKRSSVQSVAVPTSVTVTGK 148
DB 31 RTEKHSFMPDSDPVVKVTKQSRLTP-----PTMPPTTNG-APRTSSFTPT--- 74
OY 149 PLNTVTVTKAPSSL-CASSTP---SNEPRLKAENSAVAQINISPTM-LENVKKCKKFLAML 203
DB 75 TLTKGTSHSPALNGAPPPNGFSGSPSSSSSSSIANO-QLPPACGARQLSKLRFETTL 133

OY 204 IKLACSGSPSEMGONVKVLEQDLDAKTEAEFEFRKLYVELKSSPQHLVFLKSYVA 263
DB 134 QOF--GNISPEIGRVRTIVGLVNSLTITFEFRSKQDEATNPLRFVIFPLKANLPL 191
OY 264 L-ROLLPNSQSFQ---OCVOQTSSDMVIACTTTVTSPV 301
DB 192 LQRELLHCARLAKQNPQAYLAHQELLDASTSPVDSSELL 233

RESULT 13
ID YK82_YEAST STANDARD; PRT; 1169 AA.
AC P36170;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 122.2 kDa protein in STR1 3'region precursor.
GN YKRI02W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP Gaillon L., Dujon B.;
RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: STRONG, TO YEAST PROTEIN PLO1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z28327; CAAB2182.1; JOINED.
DR PIR: S38181; S38181.
DR SGP: S0001810; YKRI02W.
KW Hypothetical protein; Glycoprotein; Repeat; Cell wall; signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 1169 HYPOHETICAL PROTEIN YKRI02W.
FT DOMAIN 122 122 SER/THR-RICH.
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 656 656 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 686 686 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 1169 AA; 122164 MW; 940C6DB3369C669 CRC64;

Query Match 8.5%; Score 151; DB 1; Length 1169;
Best Local Similarity 21.2%; Pred. No. 0.067;
Matches 87; Conservative 69; Mismatches 137; Indels 118; Gaps 18;

OY 4 VTKAVPVSNAPRVSSGPRLPAPQIYAVAKAPNTTQF-PANIQLPQGVILINSNGPLML 62
DB 669 VTSSSVSTPITSESSSSASVTIL---PSTITSEFRPTMK---TKVVISSSPTNL 740
OY 63 VSPQOQTVRAETTSNITSNPAY-----PAPQTVKICTVPSNSOLIKKVAATPV 112
DB 741 ITSVTTSKDSIVGSSTSVSLSSISLPSSTVSSASSEQIFHSIVSSNQALITSSSKTV 800
OY 113 KK-----LAQIGTVVTVPR-----PSSVQSVAVPTSVTVTPGPK 149

Db 801 SSESSESHRTPTTSSSEGIKSSGVIEBTSFPHHTSTASTSVQSSQFVTPSP 860
 QY 150 LNTVTTLKSSLG----ASSTPSNEPNLKAENSAVOINLSPTMLENVKCKNFLAMLIK 205
 Db 861 ISVVA---FRSTGLNQSQTSTNSKKTMSSENSASV-----893
 QY 206 LACSSGSPQMGNKKVLBQLLDAKIEAE---FTR---KIYVELKSPQHLVPELK 258
 Db 894 MPSSSTSPKRG-----KVTSEDESSGFSRDTTYRMTSETPTN---EQ 936
 QY 259 KSVVALROLPLPNSQSF100CVQOTSSDMVATCTTV-----TT-----SPVTTTVSS 307
 Db 937 TLLITVSSCESNSCS-----NTVSSAVSTATTITTINGITTYTTCPLSATLTLTVSK 989
 QY 308 SOS-EKSIIVSGATAPRTVSVOTLNP--LAGPYGAKAGVTLHSGPPTAAT 355
 Db 990 LSESEKTLITVYSCSEGVCSFETASPAIVATATVADVTVVSTWSPQAT 1040

RESULT 14

CPN_DROME STANDARD; PRT; 865 AA.

ID CPN_DROME
 AC 002910;
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Calphotin.
 GN CPN OR CAP.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Canton-S;
 RX MEDLINE=93165729; PubMed=8094559;
 RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
 RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Canton-S;
 RX MEDLINE=93165730; PubMed=8434015;
 RA Ballinger D.G., Xue N., Harshman K.D.;
 RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
 calcium and contains a leucine zipper.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
 CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
 REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
 OF CA+2 PER MOL OF PROTEIN.
 CC -1- SUBUNIT: HOMODIMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
 CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
 COMPOUND EYES AND OCELLI.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
 DEVELOPMENT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: L02111; AAA28405.1;
 DR EMBL: L05080; AAA28420.1;
 DR PIR: A47282; A47282.
 DR Flybase: FBgn010218; Cpn.
 KW Calcium-binding.
 FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).

FT CONFLICT 43 43 I -> T (IN REF. 2).
 FT CONFLICT 64 64 I -> A (IN REF. 2).
 FT CONFLICT 76 76 T -> A (IN REF. 2).
 FT CONFLICT 100 100 P -> PP (IN REF. 2).
 FT CONFLICT 126 127 VQ -> AP (IN REF. 2).
 FT CONFLICT 154 154 I -> V (IN REF. 2).
 FT CONFLICT 160 160 S -> T (IN REF. 2).
 FT CONFLICT 534 534 A -> E (IN REF. 2).
 FT CONFLICT 699 699 I -> T (IN REF. 2).
 FT CONFLICT 703 703 V -> L (IN REF. 2).
 FT CONFLICT 721 721 D -> E (IN REF. 2).
 SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;

Query Match 8.5%; Score 150; DB 1; Length 865;
 Best Local Similarity 22.0%; Pred. No. 0.05;
 Matches 90; Conservative 59; Mismatches 176; Indels 84; Gaps 16;

QY 1 GTLVTKV-APVSA-----PKVSSCPRLPAP-----QIYAVKAPNTTIOFANLQ 45
 Db 4 GILPSVSAVPAAPVTPSAVAAPOVSPAAVAPAPAPAPAPAPAPAPAPAPAPAPAP 63
 QY 46 LPDGLVLIKNSGPIMLVSPQOTVTRAETTSNITSRPAPANQVAKICTVPNS-SSQLI 104
 Db 64 IPAPAPIAASVTPVSAVAP--PYVAAPTP-----PAASVSTPVAVAOIPVAVSAPVA 115
 QY 105 KKAIVTPVKKL-AQIGTVTVTVPKRSSVOSVAVPPISVTV-VTPGKPLNTVTLKPSLG 162
 Db 116 PVAAPTPPVQOIPVAPVATPVPVAAAPTPAVTPVIVISVIAVPPVAVPANTTVPAAP 175
 QY 163 ASSTPSNEPNLKAENSAVOINLSPTMLE-----NKKCKNFLAMLIKACS 209
 Db 176 VAAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 227
 QY 210 GSQSPENGVNKKLVEQLDAKIEAEFTKRLVLELKSSPQHLVPELKSVVALROLPL 269
 Db 228 ---PEVSVAATKPLAAAEVVAAPATERPVVAPAAASHVAVPAVETAVVA-----P 278
 QY 270 NSQSFIOCCVQOTSSDMVATCTTVTTTSPVTVTVVSSQSEKSIIV---SCATAPRV 325
 Db 279 VSAS-----TEPPVAAATLTTADET-PALAPVVAESQVAAVTAVTPPTPAPEPPT 329
 QY 326 S-----VQTLNPLAGKAGAGVTVLHSGVPTAATGTT 359
 Db 330 APPVVAETPVAASVAAVETTPVVPVPAABSTAPVAVATTVPATLAVT 378

RESULT 15

MTG8_MOUSE STANDARD; PRT; 577 AA.

ID MTG8_MOUSE
 AC 061909;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE MTG8 protein.
 GN CBPAZT1 OR CBPAZT1H OR MTG8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ICR; TISSUE-Brain;
 RX MEDLINE=96121389; PubMed=8575770;
 RA Niwa-Kawakita M., Miyoshi H., Gotoh O., Matsushima Y., Nishimura M.,
 RA Shiba H., Ohki M.;
 RT "Cloning and gene mapping of the mouse homologue of the CBPAZT1 gene
 associated with human acute myeloid leukemia.";
 RL Genomics 29:755-759(1995).
 CC -1- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: CONTAINS 1 MYND-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D32007; BAA06774.1; -
DR MGD; MG1:104793; Cbfa2l1h.
DR InterPro: IPR003894; TAF_hom.
DR InterPro: IPR002893; Znfx_MYND.
DR Pfam: PF01753; zf-MYND; 1.
DR SMART; SM00549; TAFH; 1.
KW Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 30 POLY-PRO.
FT DOMAIN 75 81 POLY-SER.
FT DOMAIN 263 270 POLY-PRO.
FT DOMAIN 380 386 POLY-SER.
FT ZN_FING 488 MYND-TYPE.
SQ SEQUENCE 577 AA; 64337 MW; 2656F185318C4B11 CRC64;

Query Match

Best Local Similarity 29.3%; Score 147; DB 1; Length 577;
Matches 65; Conservative 39; Mismatches 90; Indels 28; Gaps 10;

OY 89 QTVKICTVPNNSQLIKKVAVTPEYKLAQIGTIVVTTPKPSYQSYAVPTSVVTVTPGK 148
DB 4 RTEKSTMPDSDPVVKQTSRLTP-----PAMPPEPTTQG-APRTSFTPT--- 47
OY 149 PLNTVTTLKPSL-GASTP---SNEPNLKAENSAVOINLSPTM-LENVKKCKNFLAML 203
DB 48 TLNGTSHSPALNGAPSPGPFNGSPSSSSSSSIANO-QLPPACGAROLSKLKFLLTL 106
OY 204 IKLACSGSOSPEMGONKLIYEOLDAKIEAEFTFRKLYELKSSPOPHIVPFLKSYVA 263
DB 107 QQF--GNDISPEIGERYRTLVGLVNSTLTIEEFHSLQDEATNPELRPEVIFLKANLPL 164
OY 264 L-RQLLPNSQSFIO--QCVQOTSSDMVIACTTTVTTSPPV 301
DB 165 LQRELHCARLAKQNPQVLAQHQLLLDASTSPVDSSELL 206

Search completed: February 16, 2003, 21:56:21
Job time : 15.9186 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 25.2146 Seconds
(Without alignments)
2933.659 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Sequence: 1 GLVYTKVAPVSAAPKVS...KAGVYTHSVGPATAGTGT 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	497	28.0	662	11	091W6 mus musculus
2	194.5	11.0	851	5	08T9E0 drosophila
3	182	10.3	1795	5	076894 drosophila
4	181.5	10.2	1275	5	076602 caenorhabdi
5	181	10.2	5374	11	099ND0 mus musculu
6	177.5	10.0	1306	11	09JX31 mus musculu
7	175.5	9.9	709	4	09NVJ9 mus sapien
8	174.5	9.8	1979	11	090X35 mus musculu
9	173.5	9.8	528	6	029071 sus scrofa
10	172.5	9.7	1349	4	08WQ4 homo sapien
11	171.5	9.6	2045	11	09N4S7 caenorhabdi
12	169.5	9.6	2045	11	061191 mus musculu
13	169.5	9.6	2045	11	09QW42 mus musculu
14	167.5	9.4	549	5	061707 caenorhabdi
15	166	9.3	514	4	09YX8 homo sapien
16	165.5	9.3	2187	11	P70670 mus musculu

17	164	9.2	1322	4	08TE50 homo sapien
18	164	9.2	1893	5	09NKC9 drosophila
19	160.5	9.1	886	12	09Q887 human herpe
20	159.5	9.0	2781	4	09UC2
21	159	9.0	513	4	043418
22	158.5	8.9	1844	5	022579
23	157	8.9	847	5	019930
24	157	8.9	3178	5	096964
25	156.5	8.8	1696	4	08TEN3
26	156	8.8	645	4	09H048
27	156	8.8	998	4	012869
28	156	8.8	1057	5	09W294
29	156	8.8	1324	4	015312
30	155.5	8.8	842	5	09VGC9
31	155.5	8.8	864	5	095045
32	155.5	8.8	864	5	09VGC8
33	155.5	8.8	1151	13	057580
34	154.5	8.7	3507	5	023587
35	153.5	8.7	2232	3	0919D8
36	152.5	8.6	577	13	0919D7
37	152.5	8.6	604	13	0919D7
38	151.5	8.5	961	3	092223
39	150	8.5	873	10	09C548
40	149.5	8.4	1663	4	09UQ01
41	149.5	8.4	3261	4	09Y556
42	149.5	8.4	3664	4	096T58
43	149	8.4	586	13	091AB2
44	147	8.3	744	3	08TFC9
45	147	8.3	1029	4	08TDH7

ALIGNMENTS

RESULT 1

Q91W6 PRELIMINARY; PRT; 662 AA.

AC Q91W6; 01-DEC-2001 (TREMURel, 19, Created)

DT 01-DEC-2001 (TREMURel, 19, Last sequence update)

DT 01-JUN-2002 (TREMURel, 21, Last annotation update)

DE TANA-binding protein associated factor TAFII135 (Fragment).

GN TAF4A.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RC SEQUENCE FROM N.A.

RA STRAIN-BALB/C; TISSUE-PO BRAIN;

RT Metels M., Brunthorst A., Neuman T.;

RT "Cell Type Specific Expression of the TFIID Component TAFII135 in the

RT Nervous System."

RL Exp. Cell Res. 0:0-0(2001).

DR EMBL; AY038601; AAK94779.1; -

DR MGD; MGI:2152346; Tafa4a.

FT NON TER 1

SEQUENCE 662 AA; 71398 MW; 91A75F38CBDD0DA4 CRC64;

Query Match 28.0%; Score 497; DB 11; Length 662;

Best Local Similarity 39.6%; Pred. No. 6.6e-26;

Matches 133; Conservative 53; Mismatches 102; Indels 48; Gaps 14;

QY	20	PRLRAPQVAVKARNTTIOFPANIQLPKGVILKSNGLMYSPOQTV-----TRAE	73
DB	10	PRLRAPQVAVKARNTTIOFPANIQLPKGVILKSNGLMYSPOQTV-----TRAE	59
QY	74	TTSNITSRPAVPAPQVYKIVTSSSOLIKKAVTPVKTLAIGTVVTVTPKPSVQ	133
DB	60	QGSIMARPARPTGAPRPQISTVQAPGPITAR-QVTP-----TTIIVQV---SQAQ	107
QY	134	SVAVPTSVVTVTPKPLNTVT--TLKPSGLASS-----TPSNE--PKIAENSAVQINTL	185

Db 108 TTVOPTTLORSPGVQPOLVIGGSAQPASLGATAVOTGTPORTVPGASTSTATE--- 164
 QY 186 SEPMLENVKCKNFAMLIKACSGSQSPMGONVKLVEQDLDAIEEFTPKRYLVEL 245
 Db 165 ---TMMENVKCKSEFLTKLASSGKQSTETANVNDLVONLIDGKIEADEFSRLREL 221
 QY 246 KSSPOHLPFLFKSVVALROLLPNSOSFIQOCVOOTSSDMVATCTTTPVTTTV 305
 Db 222 NSSPQVLPFLFKRSLPALRQLTPDSAAFIQSSQOQPP---ASQATLTAIVLSSSV 277
 QY 306 SSSQSEKSIIVSGATAPRTVSQVOTLNLPLAGPYGAKA 341
 Db 278 QRTAGKTAASVTSALOPPVISL--TQPTQVGVGKQA 311

RESULT 2

Q8T9E0 PRELIMINARY; PRT; 851 AA.
 AC Q8T9E0;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE SD04735p.
 GN TAF110.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxId=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brookslein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guartin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacel J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RA Sumbit (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY069807; AAL39952.1;
 SQ SEQUENCE 851 AA; 92093 MW; 665B28B958C984C CRC64;

Query Match 11.0%; Score 194.5; DB 5; Length 851;
 Best Local Similarity 26.9%; Pred. No. 3.4e-05;
 Matches 101; Conservative 52; Mismatches 124; Indels 99; Gaps 19;

QY 22 LPAPQIVANK--APNMTTQ--FPANQLPRTGVLKNSGSLMLVSPQOTVRAETTSNI 78
 Db 93 LPAGVVGHRQAPSSQOQKMPN---PLSRVINSHAGVRPQSP--SSMTTATSN 148
 QY 79 TSRPAVPANPQTVICTVPSNSS--OLIKKAVTPVKKLAQIGTVVTPKPSVQSA 136
 Db 149 I-----VNSVASSGVANSSOPHLVQLNMQAP--QLPQI--TQITIPAOQSOQ---- 193
 QY 13 VFTSVVTVTPGKPLNTVTTTLKPSISGASSTPSNEPNLKAENSAVAOINLSPMLENVK-K 195
 Db 194 -----QOVNVSSAGCTATAVSSTTA-----ATTQOQWTKRK 226
 QY 196 CKNFLAMLIKILACSGSQSPMGONVKLVEQDLDAIEEFTPKRYLVELKSSPOHLYP 255
 Db 227 CKFLANLEL--STREPKVEKNVKTLLQELVNAVNEEPCDRERLLNLSPOQCLIG 284
 QY 256 FLKKSVALIKQL-----LPSNSQSFITQ-----QCVOQ 281
 Db 285 FLKKSPLRLRLQALYTKELVIEGIKPPQHVLAGLISQQLPKIQADIRIGSQTTTIG 344
 QY 282 TSSDMVATCTTTPVTTVSQVOTLNLPLAGPYGAKA 341
 Db 345 TQVMI--TPNALGTPRPITGHTTISKOPPN--IRLPAPRLVMTGIRIQIP--SLQVP 398
 QY 339 AKAGVIVLHVSQPTAA 354
 Db 399 GOANIVQIR--GPOHA 412

RESULT 3

Q76894 PRELIMINARY; PRT; 1795 AA.
 AC Q76894;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE EG:5667.1 protein.
 GN EG:5667.1 OR CG14796.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxId=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherzer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brookslein P., Brotler P.,
 RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Donnes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostli M., Houston K.A., Howland T.J., Mel M.-H., Ibergam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kul P., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacel J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN (2)
 RP SEQUENCE FROM N.A.
 RA Cadieu E., Dreano S., Lelaure V., Mottier S., Galbert F.,
 RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
 RL Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RA Benos P.;
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003421; AAF45644.1; -;
 DR EMBL; AL031028; CAA19845.2; -;
 DR FlyBase; FBgn0025390; EG:5667.1.
 DR InterPro; IPR002357; Chitin_bind_Pera.
 DR InterPro; IPR002365; P_rich_extensan.
 DR Pfam; PF01607; CBM_14; 2.
 DR PRINTS; PR01217; PRICHEXTENS.
 DR SMART; SM00494; CHIBD2; 2.

SO SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 10.3%; Score 182; DB 5; Length 1795;
 Best Local Similarity 22.1%; Pred. No. 0.00058;
 Matches 86; Conservative 52; Mismatches 165; Indels 86; Gaps 13;

OY 9 PVSAPPVSSGPRLLPAPQIVAVKAPNTTIOFPANLQI-----PGTVLIKNSNG 58
 DB 700 PTTSGTKRTPPK--PSTRTPPTTKVPTTQITTTPLRSSSTETSTQPPPTTTPQPTT 758
 OY 59 PLMLVSPQIVTRAFETNSITSPAVPANPQTVKICVPPNSSQILKKAIVAVPVKLAQI 118
 DB 759 TLLVTPKTSITTTTTEKPISSPK-PTTQOKTSTIAPNT-----KVALITQETPT 812
 OY 119 GTT-----VVTVPKPS-----VOSVAVPTSVTVTPKPLNT 152
 DB 813 QSTGTITTRKTTNNPPTSTKPISTTPPKPTTTPKSTVASTKTKTISPKPTTE 872
 OY 153 VTLKPSLSGASTPSPNEPNLKAENSAVAOINLSPMLLENVKKCNFLAMLIKLAGSGSQ 212
 DB 873 KSTENPPTNSVKTSALTSSQRA-----TSVTSEPTKQNTTITTPPKPTTLKTS 922
 OY 213 SPEGONKVKIYEDQLDAK--IEAEFTKRLVLEKSSQPHLVFLKKSVALROLPLN 270
 DB 923 TQETATSTQKSVYTTTTRKATESSPLTTLSTEEPNTPKP-----LKTTPPT 970
 OY 271 SQSFIOCVQOITSDMVATCTTWTTPVTVTVTVSSOSEKSIIVSGA--TAPR--TVS 326
 DB 971 TTS-----VTAITRITTTTISESSTETSTQKRSITPTTTRTPKVTYVI 1017
 OY 327 VOTNLPLAGPVGAKAGVTVLHSGPFAAT 355
 DB 1018 VSTONPPT--TTSKTSVTITTPNPSPT 1044

RESULT 4
 ID 076602 PRELIMINARY; PRT; 1275 AA.
 AC 076602;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 122.9 kDa protein.
 GN H02F09.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berts M.,
 RA Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lighting J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Geisel C., Harmon G.;
 RT "The sequence of C. elegans cosmid H02F09.";
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF077538; AAC64622.1;
 KW Hypothetical protein.
 SO SEQUENCE 1275 AA; 122924 MW; CB59B70C05959E25 CRC64;

Query Match 10.2%; Score 181.5; DB 5; Length 1275;
 Best Local Similarity 23.1%; Pred. No. 0.00042;
 Matches 93; Conservative 62; Mismatches 151; Indels 97; Gaps 16;

OY 2 TLVTAVAPVSAAPKVSGB-----RLPAPQIVAVKAPNTTIOFPANLQIPECTVIKSN- 56
 DB 275 TVTVPPTTVTGPVTVPVTVTVIPSTVSPITPTSPVTVVPSVTVVTPSTAVTKPST 334
 OY 57 --SGPLMVS--PQIVTRAFETNSITSPAVPANPQTVKICVPPNSSQILKKAIVAVPVK 113
 DB 335 VTLAPSTVTVVPSVTVVTPNTV--VTSPTVATPTTV--VTPST-----VTVPS 382
 OY 114 KLAQIGTVV-----TVPKPSSVOSVAVPTSVT-----VTPGRPLNTV 153
 DB 383 TVTVVPTTVTVTPSTVTVVAPSTV--VTVPTVMTSRSTVITPTGSSPSTAGTSLAST 440
 OY 134 TTKPSSLSGASTPSPNEPNLKAENSAVAOINLSPMLLENVKKCNFLAMLIKLAGSGSQS 213
 DB 441 AVTETISIGSSSTP-----LPSQSTLSMSLSISTYTPPS-----STAGATS 481
 OY 214 PEMGONKVKIYQQLDA--KIEAEFTKRLVLEKSSQPHLVFLKKSVALROLPLNS 271
 DB 482 PTVQOSTKPTTIGTSMSSGPTTVAPGASTESTVLQSSPTSGTIV-----TLPSG 529
 OY 272 OSFIOCVQOITSDMVATCTTWTTPVTVTVTVSSOSEKSIIVSGATAPRTVSQVTLN 331
 DB 530 SSTATAGTSPQASTVTVTVTIDISTVSGSTVTSQTASSLSSTESPTAGSSI-STVETVSSQ 588
 OY 332 P-----LAGPVGAKAGVTVLHSGPFAATGTT 359
 DB 589 PSTYIPVSSASSIVSTLSGTSSTA-----SPGTTESSGST 625

RESULT 5
 ID 099ND0 PRELIMINARY; PRT; 5374 AA.
 AC 099ND0;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ZAN (Zonadhesin).
 GN ZAN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 RX MEDLINE=21138439; PubMed=11239002;
 RA Wilson M.D., Riemer C., Martindale D.W., Schnupf P., Boright A.P.,
 RA Cheung T.D., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,
 RA Miller W., Koop B.F.;
 RT "Comparative analysis of the gene-dense ACHE/TFE2 region on human
 RT chromosome 7q22 with the orthologous region on mouse chromosome 5.";
 RL Nucleic Acids Res. 29:1352-1365(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 RA Cheung T.D., Wilson M.D., Koop B.F., Hardy D.M.;
 RT "Genomic Basis of Inter- and Intra-species Variation in Zonadhesin
 RT Domain Structure.";
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
 DR EMBL: AF312033; AAK28824.1; -;
 DR EMBL: AF046056; AAL04416.1; -;

DR MGD: MGI:106556; Zan.
 DR Interpro: IPR000561; EGF-like.
 DR Interpro: IPR003645; FOIN.
 DR Interpro: IPR000988; MAM_domain.
 DR Interpro: IPR003328; Tila_Cyrich.
 DR Interpro: IPR002919; TIL_Cyrich.
 DR Interpro: IPR001007; VME-C.
 DR Interpro: IPR001846; VME-D.
 DR Pfam: PF00629; MAM; 3.
 DR Pfam: PF01829; TIL; 25.
 DR Pfam: PF02345; TILA; 25.
 DR Pfam: PF00094; vwd; 4.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00274; FOIN; 21.
 DR SMART: SM00137; MAM; 3.
 DR SMART: SM00214; vwd; 25.
 DR SMART: SM00216; vwd; 4.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; 18.
 DR PROSITE: PS50060; MAM_2; 3.
 DR EGF-like domain; Glycoprotein.
 KW EGF-like domain; 5374 AA; 579536 MW; 90D2D8CE5DE24EB CRC64;
 SQ SEQUENCE

Query Match 10.2%; Score 181; DB 11; Length 5374;
 Best Local Similarity 25.6%; Pred. No. 0.0025;
 Matches 95; Conservative 55; Mismatches 151; Indels 70; Gaps 19;

OY 2 TLVTKVAVSAPKVVSSGRLPAPQIVAVKAPMTTITQPPANQLPPGTVLIKNSGSLM 61
 DB 852 TEVAVTVLPASIPPEET-----TPTEVTTTPEETIT--PAEVTVP-----PVS 894
 OY 62 LVSPOOTVRAETTSNITSFANVPANQYKICTVRNSSQLIKKAVTVPKLAQIGTT 121
 DB 895 IIS-EETTPTEVTVTPPEETITPAEVTVPVPSIPSEETTPTEVTVTPPEE-----TT 948
 OY 122 V--VTVPKPSSVOS--VAVPTSVTVTPGKPL--NTVTLKPSLSGASTSPNEPK 174
 DB 949 IPAEVTVP--PVSIPSEETITPEVTVTPPEETITPAEVTVPVPSI-----PSEETIT-- 1001
 OY 175 AENSAVQINLSPTMLENKKCNFLAMLIKLAGSGSQSPFEMQNVKILEOLLAKIEA 234
 DB 1002 -----PTEVTVTPPEETITPAEV-----TTPEETITPEVTVTPPASISIP 1043
 OY 235 EETRLKLYELKSSPQHLVPLFKSVVALROLPPNSQSIQOC---VQOTS---SDMT 288
 DB 1044 EE-TASLFEVTVTPPEETITPEVTVTPPEETITPAEVTVPVPSIPPEETITPPEETIT 1102
 OY 289 ACCTVTTSPTVTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVAKAGVYTLHS 348
 DB 1103 ASEETTVSQTETLLTEQSAVQTST----ACRPQPS--PLMPT--GPLLSKPPGVSMFS 1156
 OY 349 VGPPLAATVGGTT 359
 DB 1157 LAPT--TGVT 1165

RESULT 6
 O9JUK31 PRELIMINARY; PRT; 1306 AA.
 AC O9JUK31;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ATFA-associated factor.
 GN AT7IP OR AT7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Degraeve F., Bahr A., Chation B., Keding C.;
 RT "A murine Atfa-associated factor with transcriptional repressing

RT activity.";
 RL OncoGene 10:1807-1819(2000).
 DR EMBL: AJ132702; CAB7024.1; -.
 DR MGD: MGI:1858965; At7f1p.
 DR Interpro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 SQ SEQUENCE 1306 AA; 138488 MW; 1F94ABA571D29191 CRC64;

Query Match 10.0%; Score 177.5; DB 11; Length 1306;
 Best Local Similarity 25.0%; Pred. No. 0.00081;
 Matches 108; Conservative 41; Mismatches 160; Indels 123; Gaps 19;

OY 5 TKVAVSAPKVVSSGRLPAPQIVAVKAPMTTITQPPANQLPPGTVLIKNSGSLM 32
 DB 761 TSAVVSQPKLQTSATSGSLPAPLPPASTATVAVTQVPSGTPPTISLQPLVILH 830
 OY 33 -PNTTITQPPANQLPPGTVLIKNSGSLMIVSPQOTVRAETTSNITSFANVPANQY 91
 DB 821 VPVAVTSG--FQLQSHGTLVTONPSGNVEFISVQSPVTSGLTKNPSLPPL--NFKP 878
 OY 92 KICVPSNSSLIKKAVTVPKKLAQIGTTV--VTVPKPSSVOSVAVPTSVTVTPG--- 147
 DB 879 NIPSVSPSS--IQNSSTA---APLGTLAVQAVPTAHSIYQ--ARTSLPTVGPGLY 932
 OY 148 -----KPLN-----VTVTLKPSLSGASTSPNEPKLKNANQINLSPTMLENKKCN 198
 DB 933 SSSSSRGPIQMKIPISITFSPSS--SAEQNSSATPRIVAEQNTKTVDSIN-----KRAAD 986
 OY 199 FLAMLIKLAGSGS-----QSPFEMQ--QNVKLVQLLDKIEAEFRLKLYELKSS 248
 DB 987 STSOSGKASSDSGVDLMDDEESGTTDPKISPPSSVTSQPMKRPLOPILPAP 1046
 OY 249 P-OPHLVPLFKSVVALROLPPNSQSIQOCVQOTSQDMVIACTTVTSPVTTT--- 304
 DB 1047 PLQSGVPTSGPS-----QATIHVLPAPTAVTVTRPVTQVYTRL 1087
 OY 305 -YSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVAKAGVY-----PSEETIT-- 343
 DB 1088 PVPAAPAHQVYVTVLPAPTQA-----PLRGTVMAVAVQVNPONSVTVRVQTTTVY 1142
 OY 344 ----VTLHSVGP 351
 DB 1143 VNNGLTIGSAGP 1154

RESULT 7
 O9NVJ9 PRELIMINARY; PRT; 709 AA.
 AC O9NVJ9;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CDNA FLJ10688 fis, clone NT2RP300320, highly similar to partial
 DE putative p621 protein which interacts with transcription factor
 DE Spl.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
 RA Nakamura Y., Nagahari K., Masuh Y., Sasaki N.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK001350; BA91751.1; -.
 DR Interpro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 SQ SEQUENCE 709 AA; 75979 MW; F2C43814F6B241C9 CRC64;
 Query Match 9.9%; Score 175.5; DB 4; Length 709;

Best Local Similarity 25.4%; Pred. No. 0.00053;
Matches 106; Conservative 53; Mismatches 166; Indels 93; Gaps 20;

OY 3 LTKVAPVAPK-----VSSGPRLPAPQIVAVKAPNTTTI-----QFAN-----43
164 LVPVAVVSSQPKLQTVTSG-SLTAISV--LAPNTATVATTQVSGNQPISLOPL 220
OY 44 -----LQDPGTVLKNSGPIMLVSPQOTVRAETTSNITSRAVAVAN 87
221 PVILHVPVAVSQPOLQSHPGTLVNTQPSGNVEFISQSPPTVSGTLKPNVSLPSP-N 279
OY 88 PQVVKICTVNSSQILKKAIVPVKKLAQIGTIV-VTVPKBSVSVAVPTSVVTVTP 146
280 P--TKPNVSVSPSIQR--NPTASAPLGTTLAVQAVPTASHIVO-ATRTSLPVPGR 333
OY 147 GKPLNTVTLKP-----SSLGASSTPSENEPKKENSAAVOINLSPTLENVK 195
334 SGLXSPSTNGPIOMKIPISAFSTSSAAEQNSNTPRIENQNTNTIASVSKKAADTSQ 393
OY 196 CKNEFLAMLKLAGSGS-----PEMG--QNVKTLVEQLLDKIEAEFTRLKYV 243
394 CGK-----ATGSDSGVIDLTMDDESGASQDEKKNHTPTVSTMSSQPVSRLOP 444
OY 244 ELKSSP-QPHLVPLFLKSVVALKQLLPNSQSFIOCCVQOOTSMDVIACTTTVTSP--- 299
445 IQPAPPLQPSGVPPTSGPSQFTI-HLLPTAPT---TVNVRHPVTQVTRLVPRAVANH 499
OY 300 -VTTTSSSSQSEKSI--IVSGATAPRTVSQTLNPLAGPVGAKAGV---VTLHSVP 351
500 QVYTTTLPAPPAQAPLQGTWQAPAVQVNPQNSVTVRQOTTYVYVNNGLTGSTGP 557

RESULT 8
ID 090Y35 PRELIMINARY; PRT; 1979 AA.

AC 090Y35;
DT 01-MAY-2000 (TREMBREL. 13, Created)
DT 01-MAY-2000 (TREMBREL. 13, Last sequence update)
DT 01-JUN-2002 (TREMBREL. 21, Last annotation update)
DE Host cell factor CI (Fragment).
DE ABCD1 OR HCFC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platterer M., Brenner V., Reichwald K., Wiehe T., Oksche A.,
RA Rosenthal A.;
RT "Comparative sequence analysis of the mouse H1cam locus and the
RT corresponding region of human Xq28."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF133093; AAF22156.1;
DR MCD; MGI:105942; Hcfc1.
DR MCD; MGI:1349215; Abcd1.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 4.
DR SMART: SM00060; FN3; 1.
FT NON_TER 1
SQ SEQUENCE 1979 AA; 203436 MW; ABE998C5DDE688A9 CRC64;

Query Match 9.8%; Score 174.5; DB 11; Length 1979;
Best Local Similarity 23.6%; Pred. No. 0.0021;
Matches 106; Conservative 60; Mismatches 144; Indels 139; Gaps 22;

OY 10 VSAAPPKVSPPRLPAPQIVAVKAPNTTTIOPANLQDPGTVLKNSGPIMLVSPQOTV 69
471 IGSNPKQSGMAALAAAAATQKIPSSA---PVLVSPAGTIVT-----VAVTPGTTT 522
OY 70 TRAFETTSNITSRAVAVANPOT-----VACTVPSNSSO---LKKVAVTPVKKL 115

DB 523 LPA--TVKVASPVVWSPNAPATRLMLKTAAGVTSVSSAANTSTRTPIITVHKSQTVTAQ 580
OY 116 AQIGTIV-----TVPKP-----SSQSVAVPTS-----140
DB 581 AQVTTVVGVTKTITITLVKSPISVPGSALISNKGKWSVYQTRPVQTSNAGASGPV 640
OY 141 -----VTVTPGKPLNTVTT-----LKPSSLAGAST-----PSNE 170
DB 641 TQIQTGKPLPAGTILKLVTSADGKPTTITTTQASGAGKPTLIGISSVSPSTTKPGT 700
OY 171 PNLAENSAANOILSPMLENVKCKNFAML-IKLACSSQSPKMGQNVKALVEQLD 229
DB 701 TITITPMASALITQAGANGVSSPGIKSPIITITTKVTSQTAGP-----AKITAV-- 752
OY 230 AKIEAEFTRLK-VYELKSSP-QP-----HLVPFLKSVVALKQLLPNSQSFIOCCVQOTS 283
DB 753 PKITGHCQCVTOYVLKAGAGQPTILRTYP-----MGVRLVTPVTVSAVKAV----- 803
OY 284 SDVNIACCTTTVTTSPPVTTTSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
DB 804 -TTLVKGTTGVTTLGTVGTST-----LAGAGHSTSA-----SLATPITLGTI 850
OY 344 VTLHS--VGPTA-----ATGCTT 359
DB 851 ATLSQVAINPALTIVSAAGTTLTAAAGLT 879

RESULT 9
ID 029071 PRELIMINARY; PRT; 528 AA.

AC 029071;
DT 01-NOV-1996 (TREMBREL. 01, Created)
DT 01-NOV-1996 (TREMBREL. 01, Last sequence update)
DT 01-DEC-2001 (TREMBREL. 19, Last annotation update)
DE Gastric mucin (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=95275264; Pubmed=7755593;
RA Turner B.S., Bhaskar K.R., Hadzopoulos-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin."
RL Biochem. J. 308:89-96(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=94102478; Pubmed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopoulos-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Pig gastric mucin: isolation and characterization of a cDNA clone
RT with a novel tandem repeat."
RL Gastroenterology 106:200-200(1994).
DR EMBL: U10281; AAC48526.1;
FT NON_TER 1
FT NON_TER 528
SQ SEQUENCE 528 AA; 49907 MW; 0BF06879203B2BA CRC64;

Query Match 9.8%; Score 173.5; DB 6; Length 528;
Best Local Similarity 23.3%; Pred. No. 0.00051;
Matches 97; Conservative 61; Mismatches 138; Indels 121; Gaps 20;

OY 8 APVSAAPPKVSPPRLPAPQIVAVK-----APNT--TTTQFANLQLP--GTVLKS 55
108 APISSTSVQSSGSGVPTTSATSVQSSSSSAPTTTSATVQ--PSSSSSPISSTVSVQ 166
OY 56 NSGPIMLVSPQOTVTRAFETTSNITSRAVAVANPOTVACTVPSNSSQILKKAIVTPVK-- 113
DB 167 SSS-----SSAPTSATSVQSPSSSSSPISSTVSVQTSSTSTVSVQPS 215


```

Db 655 APETSTPPSSSTPVQTTTAAETTTSTPPSSSTPVQTTTAAETTTSTPPSSST 714
QY 335 GPV 337
Db 715 SPV 717

RESULT 12
Q61191
ID 061191 PRELIMINARY; PRT; 2045 AA.
AC 061191:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE C1 transcription factor C1 (HCF).
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Kristie T.M.; Dasher R.;
RT "CDNAs encoding the mouse homolog of the human transcription factor C1
RT (HCF).";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53925; AAB01163.1; -.
DR MGD: MGI:105942; Hcfcl.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 5.
DR SMART: SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

Query Match 9.68; Score 169.5; DB 11; Length 2045;
Best Local Similarity 23.48; Pred. No. 0.0048;
Matches 105; Conservative 60; Mismatches 145; Indels 139; Gaps 22;

QY 10 VSAPPKVSRRPLPAPQVAVKAPNTTIOFPANOLPPGVILKSNGLMLVSPQOTV 69
Db 537 IGSNPQSGMAALAAATAATOKIPSSA---PTAMSVAGTTIYKT-----VAATPGT 588
QY 70 TRAEITNITSRPVAVPAPNPT-----VKICTVPNSSO---LIKKVAVTPVKL 115
Db 589 LPA--TVKAVSSPVVSNPATRMLKTAAGVGTSSAANTSTRPIITVHKSGVTVAQ 646
QY 116 AQIGTIVV-----TTPKP-----SSVQSAVPTS----- 140
Db 647 AQVTTVVGVTKTITLVKSPISVPGSALISNLKVMVSVQTRKVOGSAVTAQASTGPV 706
QY 141 -----VVTTPGKPLNTVTT-----LKPSIGAST-----PSNE 170
Db 707 TQIIQTKPPLAGTILKLVTSADGKPTTIIITTTQASGAGTPTIIGISSVSPSTTKPPT 766
QY 171 PNLKAENSAVAQVNLSPMLNKKCKNFKLML-IKLACSGSGSPEMQNVKRLVEQLLD 229
Db 767 TIITIPMSALITTOAGATGVTSSPGIKSPIITITTKVMTSGTAP-----AKITTA 818
QY 230 AKIABEFTKRL-YVELKSSP-QP---HLVPFLKSVVALROLPLNSQSFIOQVOOTS 283
Db 819 PKIATGHGQCGVTOVLKAGAPGPGTILRTVP-----MGVRLVTPVVSVAVKP 869
QY 284 SDVIAICTTIVTSPVTTVSSQSEKSIIVSGATAPRTVSVOTLPLAGPVGAKAGV 343
Db 870 -TTLVVGKGTGVTGLVGTGVTSS-----LAGAGHSTSA-----SLATPITLGTI 916
QY 344 VTLHS--VGPTA-----ATGTT 359
Db 917 ATLSQVINPFAITVSAAGTTLTAAGLIT 945

```

```

RESULT 13
Q61707
ID 061707 PRELIMINARY; PRT; 2045 AA.
AC 061707:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE C1 transcription factor.
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Kristie T.M.;
RT "CDNAs encoding the mouse homolog of the human transcription factor C1
RT (HCF).";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U80821; AAD09225.1; -.
DR MGD: MGI:105942; Hcfcl.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 5.
DR SMART: SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210546 MW; 7AD38FCD78EBA9C CRC64;

Query Match 9.68; Score 169.5; DB 11; Length 2045;
Best Local Similarity 23.48; Pred. No. 0.0048;
Matches 105; Conservative 60; Mismatches 145; Indels 139; Gaps 22;

QY 10 VSAPPKVSRRPLPAPQVAVKAPNTTIOFPANOLPPGVILKSNGLMLVSPQOTV 69
Db 537 IGSNPQSGMAALAAATAATOKIPSSA---PTAMSVAGTTIYKT-----VAATPGT 588
QY 70 TRAEITNITSRPVAVPAPNPT-----VKICTVPNSSO---LIKKVAVTPVKL 115
Db 589 LPA--TVKAVSSPVVSNPATRMLKTAAGVGTSSAANTSTRPIITVHKSGVTVAQ 646
QY 116 AQIGTIVV-----TTPKP-----SSVQSAVPTS----- 140
Db 647 AQVTTVVGVTKTITLVKSPISVPGSALISNLKVMVSVQTRKVOGSAVTAQASTGPV 706
QY 141 -----VVTTPGKPLNTVTT-----LKPSIGAST-----PSNE 170
Db 707 TQIIHTRKPLAGTILKLVTSADGKPTTIIITTTQASGAGTPTIIGISSVSPSTTKPPT 766
QY 230 AKIABEFTKRL-YVELKSSP-QP---HLVPFLKSVVALROLPLNSQSFIOQVOOTS 283
Db 819 PKIATGHGQCGVTOVLKAGAPGPGTILRTVP-----MGVRLVTPVVSVAVKP 869
QY 284 SDVIAICTTIVTSPVTTVSSQSEKSIIVSGATAPRTVSVOTLPLAGPVGAKAGV 343
Db 870 -TTLVVGKGTGVTGLVGTGVTSS-----LAGAGHSTSA-----SLATPITLGTI 916
QY 344 VTLHS--VGPTA-----ATGTT 359
Db 917 ATLSQVINPFAITVSAAGTTLTAAGLIT 945

RESULT 14
Q61707
ID 061707 PRELIMINARY; PRT; 549 AA.
AC 061707:
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)

```

DE Hypothetical 60.4 kDa protein.
GN R119.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Clarke K., Bauer C., O'Neal D.;
RT "The sequence of C. elegans cosmid R119.";
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063007; AAC16427.1; -
DR InterPro; IPR004822; Histone_core.
DR InterPro; IPR003894; TAF_hom.
DR SMART; SM00549; TAFH; 1.
KM Hypothetical protein.
SQ SEQUENCE 549 AA; 60430 MW; 17AF21E0937B9B38 CRC64;
Query Match 9.4%; Score 167.5; DB 5; Length 549;
Best Local Similarity 28.1%; Pred. No. 0.0014;
Matches 63; Conservative 27; Mismatches 91; Indels 43; Gaps 8;
QY 158 PSLGASSTPSPNEPNAKNSAAVOINLSPTMLENYKCKNFAMLIKACSSQSPEM- 216
DB 102 PSMADQPGPGT-----PONSAAAAAASDD--KNYTKCVRFKLTILNT--SNNDPEMP 152
QY 217 --GQNYKKLVEQLDAKIAEAEFTKRLYEELKSSPPQHLVPLKSSVALROLPLNSQSF 274
DB 153 DKAARKYELIRGVILETAEETRLQVLYKSAQPHILPLQNTLPLARNVRNGTAS 212
QY 275 I-----GQCVOQTSDDVNIATCTVTTSPTVTTSSS 308
DB 213 VEGVNPPEGVTFNNGRTPPQPQPPQSQSQQPLEMQIIPPNQIPPMVGGPHMVS 272
QY 309 QSEKSTII-VSGATAPRTVSQTLNPLAGPVGAKAGVTLHSVCP 351
DB 273 VGAPMIRPMGPGGFSPMGLQ--GVRGPMGHQ--WVQMHPPP 312
RESULT 15
QY4X8 PRELIMINARY; PRT; 514 AA.
AC QY4X8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE P621 (Fragment).
GN 621.
OS Homo sapiens (Human).
OG Plasmid pectil.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON;
RX MEDLINE=20431278; PubMed=10976766;
RA Gunther M., Lathier M., Brison O.;

RT "A set of proteins interacting with transcription factor Spl
identified in a two-hybrid screening.";
RL Mol. Cell. Biochem. 210:131-142(2000).
DR EMBL; AJ242978; CAB45135.1; -
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENSN.
KM Plasmid.
FT NON_TER 1 1
FT 514 514
SQ SEQUENCE 514 AA; 54963 MW; D4C5E5AC598E1DBA CRC64;
Query Match 9.4%; Score 166; DB 4; Length 514;
Best Local Similarity 25.1%; Pred. No. 0.0016;
Matches 91; Conservative 41; Mismatches 132; Indels 98; Gaps 17;
QY 3 LVTKVAVSAPK-----VSSGPRLPAPQIVAVKAPNTTTI-----OPPAN----- 43
DB 181 LVTPPNAVSSQPKLQTFVTG-SLTATSV--LRPNATVYATTVQVPSGNPOTISLOPL 237
QY 44 -----LQPPGTVLIRKNSGPLMLVSPQQTVAETTSNITSRPAPVAN 87
DB 238 PVILHVPVAVSSQPOLQSHPGTLVTNQPGNVEFISVQSPPTVSGITKPNVSLPLP-N 296
QY 88 PQTAKICTVNPSSSQLIKKAAVTPVKKLAIGTTV-VTTPKSSVQSAVPTSVTVTP 146
DB 297 P--TKPNNVSPVSPSTIQ--NPTASAPLGGTLAVQAQVTAHSIVQ-ATRTSLPTVGP 350
QY 147 GKPLNTVTTLKP-----SSLGASSTPSPNEPNAKNSAAVOINLSPTMLENVK 195
DB 351 SGLYSPSTNGPQOMKIPISAFSTSSAAEQNSMTPTRIENQTKTIDASVSKRAADSTQ 410
QY 196 CKNFAMLIKACSSQS-----PEMG--QNYKKLVEQLDAKIAEAEFTKRLYV 243
DB 411 CGK-----ATGSDSGVIDLTMDDESGASQDPKRLNHTPVSTWSSSQPVSRLOP 461
QY 244 ELKSSP-QPHLVPEFKKSSVALROLPLNSQSFIOQCVOQTSDDVNIATCTVTTSPTV 302
DB 462 IQPAPPLQPSGVPTSGPSQTTI-HLLEPT-----APTIVNVTHTRVYQ 502
QY 303 TT 304
DB 503 VT 504
Search completed: February 16, 2003, 21:59:10
Job time : 33.2146 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 ; Search time 34.6501 seconds
(without alignments)
1738.213 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452
Perfect score: 2249
Sequence: 1 GLVTKVAPVSAAPRVSSGP.....VIGTPVQIKIAPGPVLSQP 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2249	100.0	801	19	AAW31494	Human hTAFII105 pr
2	2249	100.0	801	22	ABG09468	Novel human diagno
3	2249	100.0	852	21	AA157279	Transcription fact
4	521	23.2	737	15	AAW56494	TATA-binding prote
5	521	23.2	737	17	AAW06084	Human TATA-binding
6	521	23.2	737	18	AAW25019	Human TATA-binding
7	513	22.8	1023	23	AAU82954	Human homologue of
8	241.5	10.7	921	15	AAW56487	TATA-binding prote
9	241.5	10.7	921	17	AAW06077	Drosophila TATA-bi
10	241.5	10.7	921	18	AAW25028	TATA-binding prote

11	241.5	10.7	921	22	ABB61528	Drosophila melanog
12	241.5	10.7	921	22	ABB66055	Drosophila melanog
13	204	9.1	2035	15	AAW57141	Host cell factor P
14	202	9.0	1795	22	ABB69806	Drosophila melanog
15	199.5	8.9	842	22	ABB66631	Drosophila melanog
16	199.5	8.9	864	22	ABB71319	Drosophila melanog
17	195.5	8.7	5179	22	AAW24516	C899P predicted am
18	195.5	8.5	1930	22	AAW65656	Novel protein kin
19	191	8.5	2135	23	AAW21714	Human PKIN-9 prote
20	181.5	8.1	571	22	AAW27242	Human EXMAD-20 SER
21	181.5	8.1	708	22	AAW79978	Human protein SEQ
22	181.5	8.1	709	22	AAW78994	Human protein SEQ
23	181.5	8.1	709	22	AAW93063	Human protein sequ
24	181.5	8.1	709	22	AAW94334	Human protein sequ
25	181.5	8.1	881	22	AAW93811	Human polypeptide,
26	181	8.0	1296	23	ABG66702	Human novel polype
27	181	8.0	1296	23	ABG66756	Human novel polype
28	180.5	8.0	1326	23	ABG66756	Human novel polype
29	178	7.9	1127	22	AAW24260	Human ORFX ORF2414
30	177.5	7.9	662	22	AAW95541	Human protein sequ
31	176.5	7.8	1714	22	ABW71837	Drosophila melanog
32	176.5	7.8	2972	22	ABW60186	Drosophila melanog
33	176.5	7.8	3118	22	AAW50362	Human SRCAP. Homo
34	175	7.8	1658	22	ABW67620	Human SRCAP. Homo
35	173.5	7.7	557	22	AAW94078	Drosophila melanog
36	173.5	7.7	849	17	AAW06725	Human protein sequ
37	173.5	7.7	2781	21	AAW57453	FIOL protein, invo
38	173.5	7.7	2870	21	AAW95559	Human transcriptio
39	173.5	7.7	2907	21	AAW57452	Human transcriptio
40	173.5	7.7	3178	21	AAW95556	Human transcriptio
41	172	7.6	750	20	AAW05477	C. albicans Rpl1 p
42	171.5	7.6	1721	19	AAW48299	Cryptosporidium pa
43	171.5	7.6	1721	21	ABJ11727	Portion of Cryptos
44	171.5	7.6	1721	21	ABJ04045	C parvum GP900 pro
45	169.5	7.5	1057	22	ABB65440	Drosophila melanog

ALIGNMENTS

RESULT 1
AAW31494
ID AAW31494 standard; Protein: 801 AA.
XX AAW31494;
XX
XX 28-APR-1998 (first entry)
XX
XX Human hTAFII105 protein.
XX
XX
XX TADA-binding protein associated binding factor 105; human; activator;
KW hTAFII105; transcription factor; TAF10; transcriptional activation;
KW antibodies; diagnosis; therapy; biopharmaceutical industry.
XX
XX Homo sapiens.
OS
XX
XX US5710025-A.
PN
XX 20-JAN-1998.
PD
XX
XX 02-OCT-1996; 96US-0725012.
PF
XX
XX 02-OCT-1996; 96US-0725012.
PR
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Dikstein R, Tjian R;
PI
XX
XX WPI: 1998-109818/10.
DR
XX N-PSDB; AAV02872.
XX
XX DNA encoding human tata-binding protein associated factor - for
PT producing recombinant protein

XX Claim 1; Col 17-22; 12pp; English.
PS
XX This cDNA sequence represents a human tata-binding protein associated
CC factor, htraf1105, isolated from Daudi cell nuclear extracts. Tightly
CC associated subunits (TRAF's) are components of the transcription factor
CC NF- κ B and are thought to mediate transcriptional activation. This encoded
CC protein may be produced recombinantly from transformed host cells or
CC purified from human cells. htraf1105 specific binding agents such as
CC specific antibodies could be used for diagnosis (e.g. genetic
CC hybridisation screens for htraf1105 transcripts), therapy (e.g. gene
CC therapy to modulate htraf1105 gene expression) and in the
CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
CC cell specific activators or other transcriptional regulators).
XX
SQ Sequence 801 AA:
Query Match 100.0%; Score 2249; DB 19; Length 801;
Best Local Similarity 100.0%; Pred. No. 8,3e-159;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKVAVSAPPKVSQSPRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGFL 60
DB 1 GTLVTKVAVSAPPKVSQSPRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGFL 60
QY 61 MLVSPQQTIVTRAETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATPVKKLAQICT 120
DB 61 MLVSPQQTIVTRAETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATPVKKLAQICT 120
QY 121 TVVTVPRKSSVQSAVAVPTSVTPGKPLNTVTLKPSLSGASSPSPNEPMLKAENSA 180
DB 121 TVVTVPRKSSVQSAVAVPTSVTPGKPLNTVTLKPSLSGASSPSPNEPMLKAENSA 180
QY 181 VOINISPTMLENVKCKKNFLAMLIKACSGSPKMGQVKKLVQDLDAKTEAEFEFTK 240
DB 181 VOINISPTMLENVKCKKNFLAMLIKACSGSPKMGQVKKLVQDLDAKTEAEFEFTK 240
QY 241 LVELKSSPQPHLVPFLKRSVVALRQLPNSQSFIOQCVOQTSDDVAVTCTTTVTTSPV 300
DB 241 LVELKSSPQPHLVPFLKRSVVALRQLPNSQSFIOQCVOQTSDDVAVTCTTTVTTSPV 300
QY 301 VTTVSSSSQSEKSIIVSGATAPRTVSVOQLNPLAGPVGAKAGVTLHSVGPAAAGCTTA 360
DB 301 VTTVSSSSQSEKSIIVSGATAPRTVSVOQLNPLAGPVGAKAGVTLHSVGPAAAGCTTA 360
QY 361 GTGLQTSKPLVTSVAVNTVTVSLQPEKPVVSGTAVTSLPAVTEGTSGAAILCLPSVKP 420
DB 361 GTGLQTSKPLVTSVAVNTVTVSLQPEKPVVSGTAVTSLPAVTEGTSGAAILCLPSVKP 420
QY 421 VVSFCMDHICKRPVIGTPVOIKLADGPVLSOP 452
DB 421 VVSFCMDHICKRPVIGTPVOIKLADGPVLSOP 452
RESULT 2
ABG09468 ID ABG09468 standard; Protein: 801 AA.
AC ABC09468;
DT 13-FEB-2002 (first entry)
DE Novel human diagnostic protein #9459.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR N-PSDB: AAS73655.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
PS Claim 20; SEQ ID No 39827; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 801 AA:
Query Match 100.0%; Score 2249; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 8,3e-159;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKVAVSAPPKVSQSPRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGFL 60
DB 1 GTLVTKVAVSAPPKVSQSPRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGFL 60
QY 61 MLVSPQQTIVTRAETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATPVKKLAQICT 120
DB 61 MLVSPQQTIVTRAETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATPVKKLAQICT 120
QY 121 TVVTVPRKSSVQSAVAVPTSVTPGKPLNTVTLKPSLSGASSPSPNEPMLKAENSA 180
DB 121 TVVTVPRKSSVQSAVAVPTSVTPGKPLNTVTLKPSLSGASSPSPNEPMLKAENSA 180
QY 181 VOINISPTMLENVKCKKNFLAMLIKACSGSPKMGQVKKLVQDLDAKTEAEFEFTK 240
DB 181 VOINISPTMLENVKCKKNFLAMLIKACSGSPKMGQVKKLVQDLDAKTEAEFEFTK 240
QY 241 LVELKSSPQPHLVPFLKRSVVALRQLPNSQSFIOQCVOQTSDDVAVTCTTTVTTSPV 300
DB 241 LVELKSSPQPHLVPFLKRSVVALRQLPNSQSFIOQCVOQTSDDVAVTCTTTVTTSPV 300
QY 301 VTTVSSSSQSEKSIIVSGATAPRTVSVOQLNPLAGPVGAKAGVTLHSVGPAAAGCTTA 360
DB 301 VTTVSSSSQSEKSIIVSGATAPRTVSVOQLNPLAGPVGAKAGVTLHSVGPAAAGCTTA 360
QY 361 GTGLQTSKPLVTSVAVNTVTVSLQPEKPVVSGTAVTSLPAVTEGTSGAAILCLPSVKP 420
DB 361 GTGLQTSKPLVTSVAVNTVTVSLQPEKPVVSGTAVTSLPAVTEGTSGAAILCLPSVKP 420

```

Db 361 GTGLQTSKPLVTSVANTVTVTSIQPEKPVVSGTAVTSLPVAVTFGETSGAATCLPSVKP 420
QY 421 VVSECFMDHICKPVIGTVPQIKLAQPGVLSQP 452
Db 421 VVSECFMDHICKPVIGTVPQIKLAQPGVLSQP 452

RESULT 3
AAV57279 standard; Protein; 852 AA.
ID AAV57279:
XX
AC AAV57279:
XX
DE 06-JUN-2000 (first entry)
XX
DE Transcription factor subunit TAFII105 polypeptide.
XX
KW TATA box-binding protein associated factor II 105; TAFII105; cancer;
KW transcription factor; apoptosis; cytosolic; immunosuppressive;
KW antiinflammatory; virucide; antibacterial.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 802 /label= Glx
FT /note= "encoded by TGA"
FT Misc-difference 834 /label= Glx
FT /note= "encoded by TAA"
XX
PN WO200012699-A1.
XX
PD 09-MAR-2000.
XX
PF 25-AUG-1999; 99WO-IL00464.
XX
PR 27-AUG-1998; 98IL-0125971.
XX
PA (YEDA ) YEDA RES & DEV CO LTD.
XX
PI Dikstein R, Yamit-hezi A;
XX
DR WPI: 2000-256640/22.
DR N-PSDB; AA290465.
XX
PT Polypeptide encoding TATA box binding protein associated factor II 105
PT useful for treating e.g. cancers and inducing apoptosis has a dominant
PT negative effect on the normal biological activity of the binding
PT protein -
XX
PS Claim 7: Fig 2: 48pp; English.
XX
CC This represents a polypeptide comprising a (modified) fragment (I) of
CC a TATA box-binding protein associated factor II 105 (TAFII105). A
CC pharmaceutical composition comprising (I) or the polynucleotide or an
CC inhibitor or antagonist of (I) is useful for treating cancers and
CC inducing apoptosis in pathological cells. The composition is also useful
CC for treating autoimmune diseases, inflammatory processes and viral or
CC bacterial infections.
XX
SQ Sequence 852 AA;

Query Match 100.0%; Score 2249; DB 21: Length 852;
Best Local Similarity 100.0%; Pred. No. 9e-159;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKAVPVSAPPKSSGRLPAQIVAVKAPMTTIIQEPANLQLPQGVTLKSNSSGPL 60
DB 1 GTLVTKAVPVSAPPKSSGRLPAQIVAVKAPMTTIIQEPANLQLPQGVTLKSNSSGPL 60
QY 61 MLVSPQQTVAETTSNITSRPANPQVYKICTVPSNSSQLKKAVTPYKLAQIGT 120

```

```

Db 61 MLVSPQQTVAETTSNITSRPANPQVYKICTVPSNSSQLKKAVTPYKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAAVPTSVVTVTPGKPLNTVTTLKPSISGASTPSENEPMLKAENSA 180
Db 121 TVVTVTPKSSVQSVAAVPTSVVTVTPGKPLNTVTTLKPSISGASTPSENEPMLKAENSA 180
QY 181 VQINLSPTMLENVKCKNFMLAMLIKACSSQSQSPMGONVKKLVEOLLDAKIEAEETPK 240
Db 181 VQINLSPTMLENVKCKNFMLAMLIKACSSQSQSPMGONVKKLVEOLLDAKIEAEETPK 240
QY 241 LVELKSSPQPHLVPELKKSVVALROLPLNSQSFQQCVQQTSSDMVATCTTIVTSPV 300
Db 241 LVELKSSPQPHLVPELKKSVVALROLPLNSQSFQQCVQQTSSDMVATCTTIVTSPV 300
QY 301 VTTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVAKAGVYTLHSGPTAATGTTA 360
Db 301 VTTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVAKAGVYTLHSGPTAATGTTA 360
QY 361 GTGLQTSKPLVTSVANTVTVTSIQPEKPVVSGTAVTSLPVAVTFGETSGAATCLPSVKP 420
Db 361 GTGLQTSKPLVTSVANTVTVTSIQPEKPVVSGTAVTSLPVAVTFGETSGAATCLPSVKP 420
QY 421 VVSECFMDHICKPVIGTVPQIKLAQPGVLSQP 452
Db 421 VVSECFMDHICKPVIGTVPQIKLAQPGVLSQP 452

RESULT 4
AAR56494 standard; Protein; 737 AA.
ID AAR56494:
XX
AC AAR56494:
XX
DE 23-MAR-1995 (first entry)
XX
DE TATA-binding protein-associated factor hTAFII130.
XX
KW TATA-binding protein associated factor; hTAFII130; screening;
KW diagnostic; therapeutic; gene transcription regulation.
XX
OS Homo sapiens.
XX
PN WO9417087-A.
XX
PD 04-AUG-1994.
XX
PF 28-JAN-1994; 94WO-US01114.
XX
PR 28-JAN-1993; 93US-0013412.
PR 30-JUN-1993; 93US-0087119.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
DR WPI: 1994-264019/32.
DR N-PSDB; AAQ70731.
XX
PT TATA-binding protein associated protein factors - and
PT corresponding nucleotide sequence and deriv. antibodies, useful
PT in screening, diagnostics and therapeutics
XX
PS Disclosure; Page 142; 180pp; English.
XX
CC The TATA-binding protein associated factor hTAFII130 (including
CC specific antibodies and fusion products) are used in drug screening,
CC diagnostics and therapeutics. They are used in the development of
CC specific biochemical assays for screening compounds that agonise or
CC antagonise selected transcription factors involved in regulating
CC gene expression associated with human pathology.
XX
SQ Sequence 737 AA;

```

Query Match 23.2%; Score 521; DB 15; Length 737;
 Best Local Similarity 34.0%; Pred. No. 3.2e-30;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

13 PKVSSG-----PRLPAPQIVAVKAPMTTITQFPANILQPPGVILIKNSGPI 61
 68 PRATTSIGIRATLTPVLAAPRLPQ-----PQNPNTNQ---NQQLPPGWLVSSENGQL 118
 62 LVSPOQTVTR-----AETTSNITSRPVAPNPQVYKICTVPNSSQLIKKVAATPVYKL 115
 119 MI-POQALAQMOQAHAQOQTWAPRPAPTSAAPVOISTVQAPGPITLIR-QVTP----- 172
 116 AQGTGVTVTVPKSSVQSAVPTSVTVTPGKPLNTVT--TLKPSLIGSS-----TPS 168
 173 ----TITIKOV---SQAQTTVQPSATLQNSPGVQPOLVIGGAQOTSLGTAATVQGTGPQ 225
 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILACSGSSQSPSEMQUYKILVEOL 227
 226 RIVPGATTTSSATE-----TMEVKKCKNFSTLIKILACSGKQSTETPAANKVELVONL 279
 228 LPAKIAEETRLKLYELKSSPOPHLVPLKKSVALRQLPNSQSTIQOCVQO---TS 283
 280 LDKIEADEFTSRILYRELNSSPPQVLPVPLKRSIPALROLTPDPSAAFIQSQOQPPPTTS 339
 284 SDMWIATCTTVTSPVVTTVSSQSEKSIIVSGAAPTPTVSQVQTLNPLAGVGAAGV 343
 340 Q-----ATTALITAVLVSSSVQRTGKTAATVTSALQPVLSL----- 376
 344 VTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLP 403
 377 -----TQPTGVGKQGOPTPLVQ-----QPPKP-----GALINRPQV 410
 404 TEGETSGAALCLPSKPVVSFCMDHICKPVYIGPVQIKL--AOPGPVL 449
 411 TLTQT-----PMVALRQPH-NRIMLTTPQOIPLPQVPPV 446

RESULT 5
 AAW06084
 ID AAW06084 standard; Protein: 737 AA.
 XX AAW06084;
 AC
 XX
 DT 27-JAN-1997 (first entry)
 XX
 DE Human TATA-binding protein associated factor hTAFII130 protein.
 XX
 KM Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
 KM RNA polymerase II; transcription; messenger RNA; nuclear fraction;
 KM lambda-gt11; expression library.
 XX
 OS Homo sapiens.
 XX
 PN US5534410-A.
 XX
 PD 09-JUL-1996.
 XX
 PF 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-0186582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR WPT. 1996-333245/33.
 DR N-PSDB; AAT42217.
 XX
 PT Screen for cpds. that bind human TATA-binding protein associated

PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 XX
 PS Examples; Column 105-112; 86pp; English.

CC This is the amino acid sequence of the human TATA-binding protein (TBP)
 CC associated factor (TAF) designated TAFII130. The protein is a component
 CC of the TFIID fraction required for reconstituting RNA polymerase II in
 CC vitro transcription activity. The encoded protein has an estimated mol.
 CC wt. of 130 kD by SDS-PAGE.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIID, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.

XX
 SQ Sequence 737 AA:

Query Match 23.2%; Score 521; DB 17; Length 737;
 Best Local Similarity 34.0%; Pred. No. 3.2e-30;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

13 PKVSSG-----PRLPAPQIVAVKAPMTTITQFPANILQPPGVILIKNSGPI 61
 68 PRATTSIGIRATLTPVLAAPRLPQ-----PQNPNTNQ---NQQLPPGWLVSSENGQL 118
 62 LVSPOQTVTR-----AETTSNITSRPVAPNPQVYKICTVPNSSQLIKKVAATPVYKL 115
 119 MI-POQALAQMOQAHAQOQTWAPRPAPTSAAPVOISTVQAPGPITLIR-QVTP----- 172
 116 AQGTGVTVTVPKSSVQSAVPTSVTVTPGKPLNTVT--TLKPSLIGSS-----TPS 168
 173 ----TITIKOV---SQAQTTVQPSATLQNSPGVQPOLVIGGAQOTSLGTAATVQGTGPQ 225
 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILACSGSSQSPSEMQUYKILVEOL 227
 226 RIVPGATTTSSATE-----TMEVKKCKNFSTLIKILACSGKQSTETPAANKVELVONL 279
 228 LPAKIAEETRLKLYELKSSPOPHLVPLKKSVALRQLPNSQSTIQOCVQO---TS 283
 280 LDKIEADEFTSRILYRELNSSPPQVLPVPLKRSIPALROLTPDPSAAFIQSQOQPPPTTS 339
 284 SDMWIATCTTVTSPVVTTVSSQSEKSIIVSGAAPTPTVSQVQTLNPLAGVGAAGV 343
 340 Q-----ATTALITAVLVSSSVQRTGKTAATVTSALQPVLSL----- 376
 344 VTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLP 403
 377 -----TQPTGVGKQGOPTPLVQ-----QPPKP-----GALINRPQV 410
 404 TEGETSGAALCLPSKPVVSFCMDHICKPVYIGPVQIKL--AOPGPVL 449
 411 TLTQT-----PMVALRQPH-NRIMLTTPQOIPLPQVPPV 446

RESULT 6
 AAW25019
 ID AAW25019 standard; Protein: 737 AA.
 XX AAW25019;
 AC
 XX
 DT 08-OCT-1997 (first entry)
 XX
 DE TATA-binding protein associated factor, hTAFII130.
 XX
 KM TATA-binding protein associated factor; TAF; nuclear protein;
 KM RNA polymerase transcription; TATA-binding protein; TBP;
 KM Initiation.


```

QY 13 PKVSSG-----PRLPAQIVAVKAPNTTIOFPANILQIPGTVLIXNSGRLM 61
DB 414 PNTTISRATLPTTUALPRLPQ-----PQNTNIO-----NEFLPGGMVLVRENGQL 464
QY 62 LVSPQQTIVR-----AETTSNITSRPVAPNPQTVKICITVPSNSOLIKKVAATPVKLT 115
DB 465 MI-PQQAQAQMAQMAQPOPTMAFRAPPTSAAPVQISIVQAGCTILAR-QVTP----- 518
QY 116 AQGTIVTVTVPKPSSVQSAVAPTSVTVTPGKPLNTVT--TLKPSGLASS-----TPS 168
DB 519 -----TTIKQV-----SQAQTVQPSATLQSRPGVQPOLVLGGAQVATSLGATAVQGTGPQ 571
QY 169 NE-PNLKAENSAVQVQNLSPMLNENKCKNPLAMLIKLAQSGSOSPENKQVKKLVEOL 227
DB 572 RTYPGATTTSSAATE-----TMENNAKCKNFTSLIKLASSGKQSTETANNAKVELVQNL 625
QY 228 LDARKEAEETRLKLYELKSSPOPHLVPFLKSVVALROLPLNSQSTFOOCVQO-----TS 283
DB 626 LDGKIEADEPTSLRYELNLSPPQYLVPLFKRSLPALROLPLPDSANFIQSSQOQPPPTIS 685
QY 284 SDAVIATCTTIVTSPVYTTVSSQSEKSIYSGATAPRTVSQVTLNPLAGPVGAKAGV 343
DB 686 Q-----ATTALTAVVLSVSSVQRTAGKTATATYSALQPPVLSL----- 722
QY 344 VTLHSVGPPLAATGTTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLAPV 403
DB 723 -----TOPTQVGVKQGOPTPLVIO-----QPPK-----GALIRPPQV 756
QY 404 TFGETSGAMICLPSVKPVSEFCMDHICKPVIGTPVQIKLAQ 444
DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQCVNLSE 785

RESULT 8
AAR56487
ID AAR56487 standard; Protein; 921 AA.
AC AAR56487;
XX
XX 23-MAR-1995 (first entry)
XX
XX TATA-binding protein-associated factor dTAFII110.
XX
XX TATA-binding protein associated factor; dTAFII110; screening;
XX diagnostic; therapeutic; gene transcription regulation.
XX
XX Drosophila.
XX
XX WO9417087-A.
XX
XX 04-AUG-1994.
XX
XX 28-JAN-1994; 94WO-US01114.
XX
XX 28-JAN-1993; 93US-0013412.
XX
XX 30-JUN-1993; 93US-0087119.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Comal L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
XX Tjian R, Wang E, Weinzierl ROJ;
XX
XX MPI: 1994-264019/32.
XX
XX N-PSDB: AAO70724.
XX
XX
XX TATA-binding protein associated protein factors - and
XX PT corresponding nucleotide sequence and deriv. antibodies, useful
XX in screening, diagnostics and therapeutics
XX
XX Disclosure: Page 56-61; 180pp; English.
XX
XX The TATA-binding protein associated factor dTAFII110 (including
XX specific antibodies and fusion products) are used in drug screening,
XX

```

```

CC diagnostics and therapeutics. They are used in the development of
CC specific biochemical assays for screening compounds that agonise or
CC antagonise selected transcription factors involved in regulating
CC gene expression associated with human pathology.
XX
XX Sequence 921 AA;
XX
XX Query Match 10.7%; Score 241.5; DB 15; Length 921;
XX Best Local Similarly 24.8%; Pred. No. 2.8e-09;
XX Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;
XX
QY 25 POIYAVKAPNTTIOFPANILQIPGTVLIXNSGRLM-----VSPQQTIVRATTS 76
DB 133 PQSPSTLSLNLNTGQTPA-----LLVKTIDNGFQLLRVGTGTPPTVOTITNINSNS 184
QY 77 NITSRPVAPNPQTVKICITVPSNSQ-----LIKVAATPVKRLAQIGTVTVTVTP 127
DB 185 NTSTTGNHPTTQ-IRLOTVPAASMTNTATSNIIYNSVASSGVANSSQPHILTOINAQ 243
QY 128 KPSSVQSAVAPTSVTVTPGKPLNTVTTLKPSGLASSSTPSNEPNLKAENSAVQVNLSP 187
DB 244 APOLPQITQITQIPAAQSOQOQVNNVSSAGGTAVAVSSITA-----ATP 287
QY 188 TMLNENK-KCKNPLAMLIKLAQSGSOSPENKQVKKLVEOLDAKIEEETRLKLYELK 246
DB 288 TQCGNTKECKRFLANIEL--STREPKVEKNVRLILOELVANVBEFCORLERLN 345
QY 247 SSPQHLVPLFKSVVALROL-----LPNSQSTFO----- 276
DB 346 ASPQPCILGFLKSLPRLRQALYTKELVIGIKRPPQHVGLAGLSQQLKIQAOIRPIG 405
QY 277 ----QCVOQTSMDNVATCTTIVTSPVYTTVSSQSEKSIYSGATAPRTVS--VQT 329
DB 406 PSQTTITIGQTVRMI--TPNALGTPRTTIGTTTISKQPN---IRLPTARLVNTGIRT 460
QY 330 LNPAGPVGAKAGVYVTLHSVGPPLAATGTTAGTGLQTSKPLVTSVANTVTVTSLOPEK 389
DB 461 QIP-SLOVPGQANIVQIR--GPOHAQLQRTGSVQIRATTRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPVTGSETSGAMICLPSVKP 420
DB 507 -----KLTVKVGQTOIKAI-TPSLMP 527

RESULT 9
AAW06077
ID AAW06077 standard; Protein; 921 AA.
AC AAW06077;
XX
XX 27-JAN-1997 (first entry)
XX
XX Drosophila TATA-binding protein associated factor dTAFII110 protein.
XX
XX Drosophila; TATA-binding protein; TBP associated factor; TFIID;
XX RNA polymerase II; transcription; messenger RNA; nuclear fraction;
XX holoenzyme; lambda-gli1; expression library.
XX
XX Drosophila melanogaster.
XX
XX US5534410-A.
XX
XX 09-JUL-1996.
XX
XX 28-JAN-1993; 93US-0013412.
XX
XX 28-JAN-1994; 94US-0186582.
XX
XX 28-JAN-1993; 93US-0013412.
XX
XX 30-JUN-1993; 93US-0087119.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Comal L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
XX

```

PI Tjian R, Wang E, Weinzierl ROJ;
 XX WPI; 1996-333245/33.
 DR N-PSDB; AAT42210.
 XX
 PT Screen for cpds. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 XX
 PS Examples; Column 27-36; 86pp; English.
 XX
 CC This is the amino acid sequence of the Drosophila TATA-binding protein
 CC (TBP) associated factor (TAF) designated TAF110. The protein is a
 CC component of the TFIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The encoded protein
 CC has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.
 CC based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.
 XX
 SO Sequence 921 AA;
 XX
 Query Match 10.7%; Score 241.5; DB 17; Length 921;
 Best Local Similarity 24.8%; Pred. No. 2.8e-09;
 Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;
 XX
 QY 25 POIYAVKAPNTTTTIOFPANLQLPPTGVYLIKNSGPLML-----VSPQOTVRAETTS 76
 DB 133 POSSITLSTANTGOTPA-----LLVKTDNGFOLLRVCTTGPPVTQTINTSNNS 184
 QY 77 NITSRPAPVAPNPTVKTCTVPSNSSQ-----LKKVAVTVKKLAQIGTVVTVTP 127
 DB 185 NTSSTNHPPTTQ--IRLQTVPAASMTNTATSNINIVASVSSGYSANSSOPHLTQUNAQ 243
 QY 128 KPSSVQSAVPTSVYVTPGKPLNTVTTLKPSGASSTPENEPLKENSAAVOINLSP 187
 DB 244 APOLPQITQIOTIPAOOSQOOOVNNSAGCTATAVSTTA-----ATT 287
 QY 188 TMLBNVK-KCKNFLAMLIKLAGSGSOSPDMGONVKKLEOLDLAKIEEFTRLIYELK 246
 DB 288 TQOGNTEKCKKFLANLIEL--STREPKPYEKNVRTLIOELVANNVPEEFCDLERILN 345
 QY 247 SSPPHLVPLFKSSVALROL-----LPPSSQSFIO--- 276
 DB 346 ASPQCLIGFLKSLPLRLQALYTKELVIEGKPPQHVIGLAGLSQOLPRLQAOINRIG 405
 QY 277 -----QCVQOTSDDVIATCTTTVTSPVTTTVSSOSEKSIIVSGATAPRTVS---VQT 329
 DB 406 PSQTTTITGOIVRMI--TPNALGTPRPITGHTTISKOPPN---IRLPTAPRLVMGCIKT 460
 QY 330 LNPILAGPVKAGAVVTLHSVGTATAGGTAGTGLQTSKPLVMSVANTVTVLSLOPEKP 389
 DB 461 QIP-SLOYPGGANIVQIR--GPOHAQLORTSSVOIRATIRP-----PNSVPTAN----- 506
 QY 390 VVSGTAATVLSLPAVTFGETSGAALCPSPVK 420
 DB 507 -----KTLAVKVGOTQIKAL-TPSLHP 527
 XX
 RESULT 10
 ID AAM25028 standard; protein; 921 AA.
 XX
 AC AAM25028;
 XX

DT 08-OCT-1997 (first entry)
 XX
 DE TATA-binding protein associated factor, dTAF110.
 XX
 XX TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP;
 KW initiation.
 XX
 OS Drosophila sp.
 XX
 PN US5637686-A.
 XX
 PD 10-JUN-1997.
 XX
 XX 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-0188582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 PR 09-MAY-1996; 96US-0646715.
 XX
 PA (REGC) UNIV CALIFORNTA.
 XX
 PI Comat L, Dynlact BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR WPI; 1997-319113/29.
 DR N-PSDB; AAT79604.
 XX
 XX Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs
 XX
 PS Example 1; Column 35-40; 86pp; English.
 XX
 CC AAM25028 represents TATA-binding protein associated factor (TAF)
 CC polypeptide, dTAF110 (mol. weight 110kD). TAF peptides derived
 CC from dTAF110 alpha, dTAF110 beta, dTAF110, dTAF110, dTAF110,
 CC dTAF110, dTAF110, and dTAF110, their human equivalents and
 CC nucleic acids encoding them, are used to modulate transcription,
 CC including transcription initiation. TAFs are nuclear proteins involved
 CC in RNA polymerase I, II and III transcription. The peptides act by
 CC binding to a different TAF, an activator, or TBP (TATA-binding protein)
 CC or competitively inhibiting association of a TAF domain with another
 CC compound, typically a protein like TBP or another TAF, an activator,
 CC or DNA.
 CC
 SO Sequence 921 AA;
 XX
 Query Match 10.7%; Score 241.5; DB 18; Length 921;
 Best Local Similarity 24.8%; Pred. No. 2.8e-09;
 Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;
 XX
 QY 25 POIYAVKAPNTTTTIOFPANLQLPPTGVYLIKNSGPLML-----VSPQOTVRAETTS 76
 DB 133 POSSITLSTANTGOTPA-----LLVKTDNGFOLLRVCTTGPPVTQTINTSNNS 184
 QY 77 NITSRPAPVAPNPTVKTCTVPSNSSQ-----LKKVAVTVKKLAQIGTVVTVTP 127
 DB 185 NTSSTNHPPTTQ--IRLQTVPAASMTNTATSNINIVASVSSGYSANSSOPHLTQUNAQ 243
 QY 128 KPSSVQSAVPTSVYVTPGKPLNTVTTLKPSGASSTPENEPLKENSAAVOINLSP 187
 DB 244 APOLPQITQIOTIPAOOSQOOOVNNSAGCTATAVSTTA-----ATT 287
 QY 188 TMLBNVK-KCKNFLAMLIKLAGSGSOSPDMGONVKKLEOLDLAKIEEFTRLIYELK 246
 DB 288 TQOGNTEKCKKFLANLIEL--STREPKPYEKNVRTLIOELVANNVPEEFCDLERILN 345
 QY 247 SSPPHLVPLFKSSVALROL-----LPPSSQSFIO--- 276
 DB 346 ASPQCLIGFLKSLPLRLQALYTKELVIEGKPPQHVIGLAGLSQOLPRLQAOINRIG 405

```

OY 277 ----OCVOQTSSDMVIATCTTIVTTSPPVTTVSSSSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGQTVQVMI--TPNALGTPRPTIGHTTISKQPPN---IRLPTAPRLVNTGIGIRT 460
OY 330 LNPAGPVGAKAGVNTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKP 389
DB 461 QIP-SLQVPGQANIVQIR--GPQHAQLQRTGVSQIRATTRP-----PNSVPTAN----- 506
OY 390 VSGTAVTISLPAVTFGETSGAICLPVSKP 420
DB 507 -----KLTAVKVGQTOIKAI-TPSLHP 527

RESULT 11
ABB61528
ID ABB61528 standard; Protein; 921 AA.
AC ABB61528;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 11376.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmacological.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX Venter JC, Adams M, Li PMD, Myers EW;
PI
XX MPI; 2001-656860/75.
DR N-PSDB; ABL05631.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 11376; 21np + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 921 AA:
SQ
Query Match 10.7%; Score 241.5; DB 22; Length 921;
Best Local Similarity 24.8%; Pred. No. 2.8e-09;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;
OY 25 PQIVAKAPNTTITQPPANLQLEPGIVLTKNSGPMLE-----VSQQTVAETTS 76
DB 133 POSPSTLTLSTLNTGQTPA-----LWKTDNGFOLLRGTGTPPTVQQTITNTSNN 184
OY 77 NITSRAVAVANPQTVKICVTPNSSQ-----LIKVAVAVPVKKIAQIGTIVTVTP 127

```

```

DB 185 NTSTTNHPPTTO-IRLQIVPAASMTNTTATSNIIIVNSVASSGIVANSSQPPHILQLOLMO 243
OY 128 KPSSVQSAVAVTSVNTVTPCKPLNTVYTLKPSISIGASSPSPNPKAENSAVQINLSP 187
DB 244 APQLPQITQITIPAQOSQGOVNNVSSAGCTADAVSSSTA-----ATT 287
OY 188 TMLBNK-KKKNFLAMIKIACSGSSQSPKGVKKVLEQDLDLAKIEAEFTKKLYEKL 246
DB 288 TQGNNTKEKCKRFLANLIEI--STREPKPVKKNVRLIDELVANANPEEFCDRLRLIN 345
OY 247 SSPQPHLPEFKKSVVALRQL-----LPNSQSFIO--- 276
DB 346 ASPQCLIGFKKSLPRLRALTKELYEGIRPPQHVGLAGLSQQLKIQALRPIG 405
OY 277 ----OCVOQTSSDMVIATCTTIVTTSPPVTTVSSSSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGQTVQVMI--TPNALGTPRPTIGHTTISKQPPN---IRLPTAPRLVNTGIGIRT 460
OY 330 LNPAGPVGAKAGVNTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKP 389
DB 461 QIP-SLQVPGQANIVQIR--GPQHAQLQRTGVSQIRATTRP-----PNSVPTAN----- 506
OY 390 VSGTAVTISLPAVTFGETSGAICLPVSKP 420
DB 507 -----KLTAVKVGQTOIKAI-TPSLHP 527

RESULT 12
ABB66055
ID ABB66055 standard; Protein; 921 AA.
AC ABB66055;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 24957.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KM pharmacological.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX Venter JC, Adams M, Li PMD, Myers EW;
PI
XX MPI; 2001-656860/75.
DR N-PSDB; ABL10158.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 24957; 21np + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABBS7737-ABBS72072).

```

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 921 AA:

Query Match 10.7%; Score 241.5; DB 22; Length 921;
Best Local Similarity 24.8%; Pred. No. 2.8e-09;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

OY 25 POIAVAVKAPMTTITQIPANQLPGCTVILINSNGPLM-----VSPQOTVRAETTS 76
DB 133 PPSSTILSLNTGQTA-----LVKIDNSGQLLRGTTGPTVOTITNTSNN 184
OY 77 NITSPAVPANPQVTKICTVPNSSQ-----LIRKVAVTPVKKLAQIGTVVTVTP 127
DB 185 NITSTTNHPTTQO-IRLOTVPAASMTNTATSNIIYNSVASSGYANSSQPHLTQUNA 243
OY 128 KPSSVGVAVPTSVYVTPKPLNTVTTLKPSLGASSTSPNEPLKANSAAVQINLSP 187
DB 244 APOLPOTQOTIIPAOOSQOQVNNVSSAGGTATAVSSTA-----ATT 287
OY 188 TMLENVK-KCKNPLAMLIKACSGSPENGQNKLVLEOLDKITEAEFTKLYELK 246
DB 288 TQCGNTEKCKRPLANIEL--STREPKPYEKVNTLIDELVANNVPEEFCRLRLN 345
OY 247 SSPQPLHPLPKSVALKOL-----LPSQSFIO--- 276
DB 346 ASPQCLIGLKSLPLQALYKELVIGIKPPQHVGLAGLSQPLQIQOIRPIG 405
OY 277 ----QCVOQTSSDMVATCTTVTTSVYVTTSSQSEKSIYSGATPRTVS---VQT 329
DB 406 PSOTTTIGQOVNMI--TPNALGTPRPTIGTITISKOPN--IRLPTAPLVMTGIR 460
OY 330 LNPAGVAGKAGVNTLHSGPTAATGATGTLQTSKPLVTSVANTYVTVLQPEK 389
DB 461 QIP-SLOVPGQANTVOIR--GPOHQAQORTGSVOIRATRP-----PNSVPTAN----- 506
OY 390 VWSGTAVTLSPAVTFCETSGAICLPSVKP 420
DB 507 -----KLTAVKVGOTQIKAI-TPSLHP 527

RESULT 13
AAR57141
ID AAR57141 standard; Protein; 2035 AA.
AC AAR57141;
XX

DT 19-MAR-1995 (first entry)
XX
DE Host cell factor protein.
XX
KM Herpes simplex virus; herpes virus; VP16; immediate early gene;
KM host cell factor; virus infection therapy; cellular protein;
KM transcription.
XX

OS Homo sapiens.

XX
XX Key Location/Qualifiers
FT Active-site 21..31
FT Active-site /note="peptide R60"
FT Active-site 168..186
FT Active-site /note="peptide R37"
FT Active-site 333..340
FT Active-site /note="peptide R52"
FT Active-site 426..449
FT Active-site /note="peptide 362"
FT Active-site 511..526
FT Active-site /note="peptide 329"
FT Active-site 578..594
FT Active-site /note="peptide 223 first sequence"
FT Active-site 594..611

FT /note="peptide R26 1st peptide"
FT 611..623
FT /note="peptide 223 2nd sequence"
FT 723..731
FT /note="peptide 318"
FT 802..813
FT /note="peptide 299"
FT 813..820
FT /note="peptide 268"
FT 836..847
FT /note="peptide R26 2nd sequence"
FT 1010..1031
FT /note="THE TNT repeat 1"
FT 1072..1093
FT /note="THE TNT repeat 2"
FT 1101..1126
FT /note="THE TNT repeat 3"
FT 1158..1183
FT /note="THE TNT repeat 4"
FT 1286..1311
FT /note="THE TNT repeat 5"
FT 1314..1339
FT /note="THE TNT repeat 6"
FT 1349..1374
FT /note="THE TNT repeat 7"
FT 1414..1439
FT /note="THE TNT repeat 8"
FT 1774..1781
FT /note="peptide 293 2nd sequence"
FT 1808..1819
FT /note="peptide 115"
FT 1819..1840
FT /note="peptide 261 1st sequence"
FT 1853..1863
FT /note="peptide 240"
FT 1901..1919
FT /note="peptide R32"
FT 1919..1930
FT /note="peptide 261 2nd sequence"

XX W09413315-A.
XX
XX 23-JUN-1994.
XX
XX 03-DEC-1993; 93WO-US11721.
XX
XX 04-DEC-1992; 92US-0989842.
XX 12-APR-1993; 93US-0046585.
XX
XX (COLD-) COLD SPRING HARBOR LAB.
XX (TULA-) TULARIK INC.
XX
XX Herr W, Lamarco K, Wilson A;
XX
XX WPI; 1994-234207/28.
XX N-PSDB; AAO69229.
XX
XX New Host Cell Factor polypeptide(s) and nucleic acid - are used
XX to develop agents for diagnosis or treatment of disease
XX associated with expression of a HCF-modulated gene e.g. viral
XX infections
XX
XX Disclosure; Page 39; 71pp; English.
XX
XX HCF is required for the transcription of a number of
XX viral genes, such as the immediate early herpes simplex virus-1
XX genes. Epitopes of the encoded protein can be used in
XX defining functional domains of HCF, identifying compounds that
XX associate with HCF or designing compounds capable of modifying HCF
XX transcription. Such agents can be used to treat viral infections.
XX
XX Sequence 2035 AA;

Query Match 9.1%; Score 204; DB 15; Length 2035;
 Best Local Similarity 23.5%; Pred. No. 5.1e-06;
 Matches 119; Conservative 66; Mismatches 188; Indels 134; Gaps 24;

10 VSAPKVVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPCTVLIKNSGMLVSPQQT 69
 537 IGSFPMSCMALAAAAATOKIPSSA---PVLVSPAGTTIVKT-----MAVTEGTT 568
 70 TRAEFTSNITSRPVAPNPQVTKICVNPSSSQLIKKVAVTPVKLAQIGTIV----- 122
 589 LPA--TVKVAASPVVMSNPAT-----RMLKTA-----AQVCTSVSSAINTS 628
 123 ---VTVPPRSSQVAVPVSIVTVTPKRLNVTVTLK-PSSL-GASSPTSNPNL----- 173
 629 TRPITVHSGV-TVAAQAQVTVVGVGTAKTITLVKSPISVPGSALISNLGKVMV 687
 174 --KAENSAVQINLSPMLNENKCKNFL--AMLIKACSGSQSP-----EMGONVK 221
 688 QIKPVQTSVNTGASTGPTQIOTGKPLPAGTILKLVTSADGKPTTITTTQASGAGTK 747
 222 KIVEQLDAKIEAEFTKLYVELKSSPOPHLVPLKKSVALROLPLNSQFTQOCVOQ 281
 748 PTLIGI-----SSVSPSTT---KPGTITIKTIPMSAITTQAGATG 785
 282 TSDMWIATCTTIVT-----SPVVT-----TVSSSGSEKSIIVSGA-----TA 321
 786 VTSFSGIKSPITITTKVMTSGTGAPAKITTAVPKIAIGHGQGVTVQVVLKGAPOGPTI 845
 322 PRT-----VSVQTLNPLAGPVGAK--AGVTLHSVGPPTAAGTGTAGTGLQTS 368
 846 LRIVPAGVRLVTPVSVAKRAVTLVVGKGTGVTTLGIVGTGET--SLAGAGHST 903
 369 KPLVISA--NTVTVSLDPEKPVSGTAVTSLPAVTEGSGAICLPVSKPVVSEFCW 426
 904 ASLAFPTITLTGATLSSO---VINPFAIVSAQTTLIAAGLTTPTITMQPV----- 954
 427 DHICKRVIPTVOIKL-AOPGVLSP 452
 955 -----SOPVQVTLITAPSGVEAP 973

RESULT 14
 ABB69806
 ID ABB69806 standard; Protein; 1795 AA.
 XX
 AC ABB69806;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 36210.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX
 OS pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PD 23-MAR-2001; 2001WO-US09231.
 XX
 PF 23-MAR-2001; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 DR N-PSDB; ABL13909.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 36210; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB161757-AB161757).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 1795 AA.

Query Match 9.0%; Score 202; DB 22; Length 1795;
 Best Local Similarity 21.8%; Pred. No. 6e-06;
 Matches 95; Conservative 58; Mismatches 196; Indels 86; Gaps 13;

9 PVAPKVVSSGPRLPAPQIVAVKAPNTTIOFPANLQ-----PPGYLISNSG 58
 700 PISSTGKPTTPK-PSTRTPPTTKVTTTQITTTPLASSTETTSQPEPTTTPQPTT 758
 59 PLMLVSPQOQTRAEFTSNITSRPVAPNPQVTKICVNPSSSQLIKKVAVTPVKLAQI 118
 759 TRLVTPKSTTTTTEKPTTSSPK-PTTQOKTSTAPNTT-----KVAITQKETTP 812
 119 GTT-----VTVTPKSS-----VQSAVPTSVTVTPGKPLNT 152
 813 QSTSTIETRTKTTNNPEPTSEKPTSTPKPSTTTPKTSVAASGTEKTTISSPKPTE 872
 153 VTLKPSLGSASTPSNEPLKKAENSAVQINLSPMLEVKKCKNFLMLIKLACSGQ 212
 873 KSTENPTNSVKTSLSSSTORA-----TSTSEPTKTONITTTTPPTLKTS 922
 213 SPKMGONVKLIVQLDAK--IEAEFTKLYVELKSSPOPHLVPLKKSVALROLPLN 270
 923 TQENATSTQKSVTVTITTKKATESPLTLTSTEBPNTTPP-----LRTTTP 970
 271 SOSFTQOCVOQISDMVIATCTTIVTTPSVVTVTVSSSGSEKSIIVSGA--TAPR--TVS 326
 971 TTS-----VTATRTITTTTISESSTETSTOKPKSTPTSPTRTPKVTYVI 1017
 327 VQTLNPLAGPVGAKAGVTLHSVGPPTAAGTGTAGTGLQTSKPLVTSVANTVTVSLQ 386
 1018 VSTONPTT--TTSKTSVITTTTPNPSSTORPTTTRPTSTASTSIGTRIPPTTNP 1075
 387 EKPVSGTAVTSLP 401
 1076 QNSTSSDLDLTVTRP 1090

RESULT 15
 ABB6631
 ID ABB6631 standard; Protein; 842 AA.
 XX
 AC ABB6631;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 26685.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX
 OS pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;

XX WPI: 2001-656860/75.

XX N-PSDB; ABL10734.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure; SEQ ID NO 26685; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABBS7737-ABBS72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pcl_sequences.

XX SQ Sequence 842 AA;

Query Match 8.9%; Score 199.5; DB 22; Length 842;

Best Local Similarity 22.5%; Pred. No. 3.3e-06; Mismatches 203; Indels 153; Gaps 25;

Matches 124; Conservative 70;

QY 1 GTLTKV-APVSAP-----PKVSSGRLPAP---QIVAKAPNTTIOFPANLQ 45

DB 4 GTIRSPVSAAPAAVPAVPAVQVSPAAPAAAPAAPIAVTPVAPPTLASVQPAIVT 63

QY 46 LPFGTVLTKSNGPLMVSPOQIVTRAETTSNITSRPAVPANPOTVKICTVPS-SSQLI 104

DB 64 IPAPAPIAASVAVASVAP--PVAAAPTTP-----AASPSTPVPVAAQIPVAVSAVA 116

QY 105 KKVAVT-----PVKKLAOIGTT--VMTTVPSSVQSAVPTSVVTGPKKELNTVTTLK 157

DB 117 PPAVATPTPAAPVPAAPVATPPVAAASAPTPPAVTPVSP--VIATPPVVPANTTV--- 171

QY 158 PSSIGASTSPSEPNLKAENSAVOINLSPTMLP-----NVKKCKNFLMLI 204

DB 172 PVAAVPAVAAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 228

QY 205 KLACSGSGSPBMGQNVKLYEQLDAKIEAEETFRKLYVELKSSPQPHLVFLKSSVAL 264

DB 229 -----PEVSVAATKPLAAAPVVAAPATETPVVAPAAASPHVSAVAVETAVVA- 278

QY 265 RQLLPNSQSFIOCCVQTSMDVIACTTIVTSPVTTTVSSSQSEKSIIV-----SGAT 320

DB 279 ----PVASAS-----TEPPVAAATLTATPET--PALAPVVAESQVAAANTVAAATPPAP 325

QY 321 APRTVS-----VOTLNPPLACPVAKAGVTLHSGVPTAATGCTAGTGL 365

DB 326 EPERIAPVVAETPEVASVAAETTPPVVPAAS-----IPAPVAT----- 369

QY 366 QTSKPLVTSVANTVTVSLQDEKP-----VSGTAVTIS-----LPAVT 404

DB 370 -TPVPATLAVDPPVTAASAVDELPPVIAISPVAVETPVDLAPVLPVAAEPVPAVV 428

QY 405 FGET-----SGAICLPVSKPVVSRCD-----HICKPVIGTP-----VOIKL 442

DB 429 AETPETPAPASAPVTTAAALDIPVAVPIAASDAPEAPSAAPIVSTPPTTASVPETT 488

QY 443 AQPGLVLSQP 452

DB 489 APAAVPTPEP 498

Search completed: February 16, 2003, 21:54:59
Job time : 42.6501 secs

This Page Blank (uspto)

Db 1 GRLVTKVAVSAPKVSQGRPLPAQIVAVKARNTTITQEPALQLEPGTVILKSNGL 60
 QY 61 MLVSPQQTVAETNTSNTSREPAVNPQVTKICIVPNSSQLIKKVAATPVKKLAQIGT 120
 Db 61 MLVSPQQTVAETNTSNTSREPAVNPQVTKICIVPNSSQLIKKVAATPVKKLAQIGT 120
 QY 121 TVYVTPKPSVQSVAVPTSVYVTPGKPLNTYTTTLKPSLSGSPNSNPINKAESNA 180
 Db 121 TVYVTPKPSVQSVAVPTSVYVTPGKPLNTYTTTLKPSLSGSPNSNPINKAESNA 180
 QY 181 VOINLSPTMLENVKCKNFKLMLIKLACSGSGSPENGMQVKKLVEOLLDAKIEEETPK 240
 Db 181 VOINLSPTMLENVKCKNFKLMLIKLACSGSGSPENGMQVKKLVEOLLDAKIEEETPK 240
 QY 241 LVEELKSSQPHLVPFLKKSVAALROLLENSGSIQCCVOQSSDMVIATCTTPTVTSV 300
 Db 241 LVEELKSSQPHLVPFLKKSVAALROLLENSGSIQCCVOQSSDMVIATCTTPTVTSV 300
 QY 301 VTTTVSSSQSEKIIYSGATAPRTVSQTLNPLAGPVGAKAGVYVTHSVGPTAATGTTA 360
 Db 301 VTTTVSSSQSEKIIYSGATAPRTVSQTLNPLAGPVGAKAGVYVTHSVGPTAATGTTA 360
 QY 361 GGGILQTSKPLVTSVANTVTVVSLQPEKPVSGTAVTLSPAVTFGETSGAATCLPSVKP 420
 Db 361 GGGILQTSKPLVTSVANTVTVVSLQPEKPVSGTAVTLSPAVTFGETSGAATCLPSVKP 420
 QY 421 VVSFCMDHICKPVYIGTPVQIKLAQPGPVLSQP 452
 Db 421 VVSFCMDHICKPVYIGTPVQIKLAQPGPVLSQP 452
 RESULT 2
 ID T2D3_HUMAN STANDARD: PRT: 1083 AA.
 AC 000268; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription Initiation factor TFIIID 135 kDa subunit (TAFII-135)
 DE (TAFII135) (TAFII-130) (TAFII130).
 GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OC NCBI-TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97336072; PubMed-912867;
 RA Mengus G., May M., Carre L., Champon P., Davidson I.;
 RT "Human TAFII135 potentiates transcriptional activation by the AF-2s
 of the retinoic acid, vitamin D3, and thyroid hormone receptors in
 mammalian cells.";
 RL Genes Dev. 11:1381-1395(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21638749; PubMed-11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 Jones M., Scavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
 Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 Leharasalo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 Marsh S.L., Martin S.L., McConachie L.J., McLean K., McMuray A.A.,
 Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Philimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Seha H.K., Showkneen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tyman A.C., Vaudin R., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE-97098442; PubMed-8942982;
 RA Tanese N., Saito D., Vassallo M.F., Chen J.-L., Admon A.;
 RT "Molecular cloning and analysis of two subunits of the human TFIIID
 complex: hTAFII130 and hTAFII100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
 CC BY THE AF-25 OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Y11354; CAA72189.1; -
 DR EMBL: AL137077; GAC36006.1; -
 DR EMBL: AL109911; CAC22312.2; -
 DR EMBL: U75308; AAC50901.1; -
 DR TRANSFAC: T02328; -
 DR GeneW: HGNC:11537; TAF4.
 DR MIM: 601796; -
 DR InterPro: IPR003894; TAF_hom.
 DR SMART: SM00549; TAFH.1.
 DR Transcription regulation; Nuclear protein.
 KM
 FT DOMAIN 39 42
 FT POLY-HIS.
 FT DOMAIN 52 57
 FT POLY-ALA.
 FT DOMAIN 98 101
 FT POLY-GLY.
 FT DOMAIN 142 148
 FT POLY-ALA.
 FT DOMAIN 266 275
 FT POLY-PRO.
 FT DOMAIN 331 337
 FT POLY-ALA.
 FT DOMAIN 680 683
 FT POLY-PRO.
 FT DOMAIN 808 813
 FT POLY-ALA.
 FT DOMAIN 828 831
 FT POLY-ASP.
 FT CONFLICT 105 117
 FT PGSPSRRLPLVPA -> GRLGLQGRGREGS
 FT (IN REF. 3)
 FT A -> S (IN REF. 2).
 FT G -> GP (IN REF. 2).
 FT MISSING (IN REF. 3).
 FT P -> L (IN REF. 3).
 FT CONFLICT 293 293
 FT SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;
 SO
 Query Match 23.2%; Score 521; DB 1; Length 1083;
 Best Local Similarity 34.0%; Pred. No. 1,1e-22;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;
 QY 13 PVRVSSG-----PLPAPQIVAVKARNTTITQEPALQLEPGTVILKSNGL 61
 Db 414 PRTVTSIGIRATLTPTVLAPELPQP-----PQNPVING--NFQDPGVLVSENGQL 464
 QY 62 LVSPQQTVAETNTSNTSREPAVNPQVTKICIVPNSSQLIKKVAATPVKKLAQIGT 115
 Db 465 MI-PQALMQMOMAHNQPTTMAPRAPRTSAPVQISIVAGPPIIAR-QVTP----- 518

QY 116 AQTGTVTVTPKSSVQAVPTSVTVTPKRLNTVT--TLKPSISLGASS-----TPS 168
 Db 519 -----TTIKOV---SQAQTAVQPSATLQSRPGVOPOLVGGAAQATATGTAATVQGTGPQ 571
 QY 169 NE-PNLKAENSAVAQVIMLSPMLNENKCKNFKLMLIKLACSSGSPSEMGNVAKLVEQL 227
 Db 572 RYTPGATTTSSAATE-----THENVKCKNFKLSTLKLKSSGQSTETANVELVQNL 625
 QY 228 LDKAEAEETRLKLYELKSSPOPHLVPFLKRSVALROLPLNSQSFLOQCVO-----TS 283
 Db 626 LDKAEAEETRLKLYELKSSPOPHLVPFLKRSVALROLPLNSQSFLOQCVO-----TS 283
 QY 284 SDVIAATCTTTTSPVYTTTSSQSEKSTYSGATAPRTVSQTLNPLAGPVAKAGV 343
 Db 686 Q-----ATTALTAVALSSSVORTACKTATATVSAQPPVLSL----- 722
 QY 344 VTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVTVSLQPEKPVVSGTAVTLSPAY 403
 Db 723 -----TOPTOVGVKGOGPTPLVIO-----QPKP-----GALIRPPQV 756
 QY 404 TFGTSCAAICLPSKPVSPFCMDHICKPYIGPVPQIKL--ADGPVYL 449
 Db 757 TLMQT-----PMVALRQPH-NRIMLTTPQIQIQLNLPQVPPV 792

RESULT 3
 ID T2D3_DROME STANDARD; PRT; 921 AA.
 AC P47825; P49845; 09YUVY7;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription initiation factor TFIID 110 kDa subunit (P110)
 GN TAF110 OR CG5444.
 OS Drosophila melanogaster (Fruit fly)
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_Taxid=7227;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 519-540; 597-616 AND 857-874.
 RC TISSUE=Embryo;
 RX MEDLINE=93115326; PubMed=7678780;
 RA Hoey T., Weinzierl R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,
 RT Tjian R.;
 RT "Molecular cloning and functional analysis of Drosophila TAF110
 RT reveal properties expected of coactivators.";
 RL Cell 72:247-260(1993).
 RN (2)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 398-406; 520-540 AND 860-877.
 RC TISSUE=Embryo;
 RX MEDLINE=93317591; PubMed=8327460;
 RA Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.;
 RT "The Drosophila 110-kDa transcription factor TFIID subunit directly
 RT interacts with the N-terminal region of the 230-kDa subunit.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agdayani A., An H.-J., Andrews-Plambeck C., Baldwin D.,
 RA Ballew R.M., Basu A.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Broksstein P., Brotler P.,
 RA Burtis K.C., Busan D.A., Butler H., Caden E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dublin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rebert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.);
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
 CC CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
 CC AND REPRESSORS. MAY FUNCTION AS A COACTIVATOR BY SERVING AS A SITE
 CC OF PROTEIN-PROTEIN CONTACT BETWEEN ACTIVATORS LIKE SPI (OR BTJ)
 CC AND TFIID COMPLEX.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L06861; -; NOT_ANNOTATED_CDS.
 DR EMBL: S63550; AAB27433.1; -
 DR EMBL: AE003528; AAF49536.1; -
 DR TRANSFAC: T02121;
 DR FLYBASE: FBgn0010280; TAF110.
 DR InterPro: IPR003894; TAF_hom.
 DR SMART: SM00549; TAFH; 1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 66 82 POLY-GLN.
 FT DOMAIN 108 111 POLY-GLN.
 FT DOMAIN 259 265 POLY-GLN.
 SQ SEQUENCE 921 AA; 99338 MW; 27E6852859872767 CRC64;

Query Match 10.7%; Score 241.5; DB 1; Length 921;
 Best Local Similarity 24.8%; Pred. No. 9.1e-07;
 Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

QY 25 POIYAVKAPNTTITQIPANLQLPQGVILKNSGPIML-----VSPOQTVRAETTS 76
 Db 133 PQSISITLSTINTQOTRA-----LLVKIDNGFOLLRVGTTGPPVYVOTITNTSINS 184
 QY 77 NITSRAVPANPQTVKICVYPNSSQ-----LIKKVAVTVPKYLAQIGTVTVTPV 127
 Db 185 NITSTTHPTTQ-IRIQTVAAASMTNTATSNIVNASVSGXANSSQPHLTQUNAQ 243
 QY 128 KPSSVQSAVPTSVTVTPKPLMTVTVTLKRSISLGASTPSENENLKAENSAVAQVIMLSP 187
 Db 244 APQLPQITQITQIPDAQSQQVQVNNVSSAGCTATAVAVSTTA-----ATT 287

QY 188 TMLNFK-KCKNFIAMLIACSGSPKMGONKVLQDLAKTEAEFTTKLYVELK 246
 DB 288 TQGNKKECKRFLANLIEI--STREPKVEKNVRLIELVNAVPEEFCDLERELN 345
 QY 247 SSFOPHLVPELKKSVYALROL-----LPNSOSFIO--- 276
 DB 346 ASBPOLCIEFLIKSLPLLNALYTKRELVEGICPPRONHVLGAGLSQOLPKIOAQIRPIC 405
 QY 277 ---QCVQOTSSDMVATCTTTVTYVSSQSEKSIIVGATAPRTVS---VOT 329
 DB 406 PSQTTTIGOTVORMI--TTPNALCTPRPTIGHITISKQPN---IRLPAPRLVNTGIRI 460
 QY 330 LNPGLGPAKAGVYTLVHVGPTAAGTGTAGGLQTSKPLVTSVANTYVTSLOPEK 389
 DB 461 QLP-SLQVPGANIVQIR--GPQHAQLQRTGVSQVQIRATRRP-----PNSVPLN----- 506
 QY 390 VVSGTAVTSLPVTGEGTSGAICLPSVYP 420
 DB 507 -----KLPAVKVGTQIRKAI--TPSLHP 527

RESULT 4
 HFC1_HUMAN STANDARD: PRT: 2035 AA.
 ID HFC1_HUMAN
 AC P51610
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF) (CEF).
 GN HFC1 OR HCE1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=93327419; PubMed=8392914;
 RA Wilson A.C., Lamarco K., Peterson M.G., Herr W.;
 RT "The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein.";
 RL Cell 74:115-125(1993).
 RN [2]
 RP SEQUENCE OF 65-2035 FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Retal brain;
 RX MEDLINE=95130107; PubMed=7829097;
 RA Fratini A., Faranda S., Redolfi E., Zucchi I., Villa A., Patrosso M.C., Strina D., Susani L., Vezoni P.;
 RT "Genomic organization of the human VP16 accessory protein, a housekeeping gene (HCE1) mapping to xq28.";
 RL Genomics 23:30-35(1994).
 RN [3]
 RP PROCESSING.
 RX MEDLINE=9603796; PubMed=7590226;
 RA Wilson A.C., Peterson M.G., Herr W.;
 RT "The HCF repeat is an unusual proteolytic cleavage signal.";
 RL Genes Dev. 9:2445-2458(1995).
 CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOIF-BINDING PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.
 CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N- AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT NONCOVALENTLY, ASSOCIATED.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A SHORT FORM, MAY BE PRODUCED BY ALTERNATIVE SPLICING. IN THE SHORTER FORM THE N- AND THE C-TERMINAL FRAGMENTS FAIL TO ASSOCIATE.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL TISSUES AND THE ADULT KIDNEY. PRESENT IN ALL TISSUES TESTED.

CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE SIGNAL.
 CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS AT A DEFINED SITE, PECE/THER, WITHIN THE HCF REPEAT.
 CC -1- PTM: GLYCOSYLATED. CONTAINS O-LINKED N-ACETYLGLUCOSAMINE RESIDUUS.
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb.ch/announce/> or send an email to license@isb.ch).
 CC EMBL: L20010; -; NOT ANNOTATED. CDS.
 DR EMBL: X79198; CA455790.1; ALT_INIT.
 DR Genew; HGNC:4839; HCF1.
 DR MIM; 300019; -
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF00041; fn3.1.
 DR Pfam; PF01344; Kelch; 5.
 DR SMART; SM00060; FN3; 1.
 DR Nucleic protein; Repeat; Alternative splicing; Glycoprotein.
 KW REPEAT 44
 FT REPEAT 93
 FT REPEAT 140
 FT REPEAT 148
 FT REPEAT 217
 FT REPEAT 265
 FT REPEAT 313
 FT DOMAIN 1010
 FT REPEAT 1010
 FT REPEAT 1072
 FT REPEAT 1101
 FT REPEAT 1158
 FT REPEAT 1286
 FT REPEAT 1311
 FT REPEAT 1314
 FT REPEAT 1349
 FT REPEAT 1414
 FT REPEAT 1439
 FT VARSPLIC 382
 FT CONFLICT 564
 FT CONFLICT 603
 FT CONFLICT 1164
 FT CONFLICT 1873
 FT SEQUENCE 2035 AA; 208841 MW; 99207FEB875204C0 CRC64;
 SQ

Query Match 9.2%; Score 206.5; DB 1; length 2035;
 Best Local Similarity 23.3%; Pred. NO. 0.00024;
 Matches 135; Conservative 71; Mismatches 194; Indels 179; Gaps 30;

QY 2 TLVTKAVPVASAPPKVSSGRLP-----APQIVAVKAPMTT-----IQ 39
 DB 446 TLDPQAP-APPPTTTIOVLPTVPSSISVPTAARTQGVPAVLKVTGPATGTGLPEVM 503
 QY 40 FPN-----LQLPPTVL-----KNSGPLMLVSPQOT-----VTRAEITSN--IS 80
 DB 504 RPAOSAGKAPVTSLPAGRVMPVPTOSAGTIGSSPQMSGMAALAAAAAATQIKPPSS 563
 QY 81 RP--AVPAN-----PQVTKICVP-----NSSSOLIKKVAVTPVKKLQ 117
 DB 564 RPLVLSVPAGTIVTKMAVTPGTTTLPAIVKAVASPPVMSNPATRMKTA-----AQ 616
 QY 118 IGTIV-----VTYKPPSSVOSAVPVSVTTPGKPLNTVYTLK-PSSL-GASS 165
 DB 617 VGSVSASATSTRPITIVKSGTV-TVAQAQVVTTVGCVKRTITLVKSPISVPGGSA 675
 QY 166 TPSENRPL-----KENSAAVOINISPTMLEVKKCKKNFL--AMLIKACSSQSD--- 214
 DB 676 LISNLGKVMVSVOYTKPQVSAVVGASTGTPVTOIITKGPLPAGTILKLTVSADGKPTTI 725
 QY 215 -----EMGQVVKVLVQDLDAKTEAEFTTKLYVELKSSQPHLVPLFKSVYALROL 269

```

Db 736 ITTQASGAGTKPILIGI-----SSVSPSTT---KPGTTIITIP 773
OY 270 NSQSFIOQVOGTSSDMVIAITCTTWT-----SPVVT-----TTVSSSQSEKSI 314
Db 774 MSALITQAGATGVTVSSGIGSPITITITTKVTSCTGAPAKITTAVPKIALIHGGQGVQV 833
OY 315 IVSGA-----TAPRT-----VSQTLNPLAGPVGAK--AGVTLHSVPTAATG 356
Db 834 VLKGAFCQPGCTILRTVMGVRVLTPTVSAVKPAVTTLVKGTGTITGTVGVST- 892
OY 357 GTTAGTGLTQSKPLVTSVA--NTVTVSLQPEKPVSGTAVTSLAVPTGEGTGAIC 414
Db 893 -SLAGAGHSTASLAPITTLTGATLSSQ---VINPATAVSAQTTTLTAAGELTP 947
OY 415 LPSVKPVSFCDHICKPVTGTPQIKL-AQGPVLSOP 452
Db 948 TITMQPV-----SQPQVTLTAPSGVEAP 973

RESULT 5
HFCL MESAU STANDARD; PRT; 2090 AA.
AC P51611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFCL) (VCAF) (CFP).
GN HCFCL.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PCE/THET, WITHIN THE HCF REPEAT (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: DA5419; BAA08258.1;
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR SMART: PF01344; Kelch; 5.
DR SMART: SM00060; FN3; 2.
KM Nuclear protein; Repeat.
FT REPEAT 44 89 KELCH 1.

```

```

FT REPEAT 93 140 KELCH 2.
FT REPEAT 148 194 KELCH 3.
FT REPEAT 217 265 KELCH 4.
FT REPEAT 266 313 KELCH 5.
FT DOMAIN 1010 1448 8 X 26 AA APPROXIMATE REPEATS.
FT REPEAT 1010 1035 HCF REPEAT 1.
FT REPEAT 1072 1097 HCF REPEAT 2.
FT REPEAT 1101 1126 HCF REPEAT 3.
FT REPEAT 1157 1182 HCF REPEAT 4 (DEGENERATE).
FT REPEAT 1295 1320 HCF REPEAT 5.
FT REPEAT 1323 1348 HCF REPEAT 6.
FT REPEAT 1358 1383 HCF REPEAT 7 (DEGENERATE).
FT REPEAT 1423 1448 HCF REPEAT 8.
SQ SEQUENCE 2090 AA; 214942 MW; E495EBB1F2385E17 CRC64;

Query Match 9.28; Score 206; DB 1; Length 2090;
Best Local Similarity 22.38; Pred. No. 0.00027;
Matches 118; Conservative 67; Mismatches 167; Indels 176; Gaps 24;

OY 10 VSAPPKVSSGRLPAPQIVAVKAPNTTIIQPPANLQIPGTVLJKNSGPIMLVSPQOTV 69
Db 537 IGSNPQSGMALALAAAAATOKIPSSA---PTVLSVAGTITVKT-----VAITPGTTT 588
OY 70 TRAEITNITSRPVAVNPQT-----VRICTVNSSSQ---LIRKVAATPVKRL 115
Db 589 LPA--TVKAVSSPVWVSNPATRMLKTAAGVTSVSAANTSTPIITVHKSQVTAQV 646
OY 116 AQIGTTV-----TVYKRP-----SSQVAVAPPS----- 140
Db 647 AQVTVVAGVTKTITLVKSPISVPGSALISNLKAVSVQTPVQTSAVTGAQSTGPV 706
OY 141 -----VTVTPGKPLNTVTTL-----RPSISGASST-----PSNE 170
Db 707 TQIIQTKPLPAGTILKLVTSADGKPTIITTTQASGAGSPITILGSSVSPSTTKPCTT 766
OY 171 PNLKENSAAVOINLSPTMLENVKCKNFLML-ITLACSGSQSPENQVAKLVEQLLD 229
Db 767 TIITIPMSALITQAGATGVTSFGIKSPIRITITTKVMTSGTAP-----AKITAV-- 818
OY 230 AKIAAEETRL-VYELKSP-QHLVPLFKSVYALQQLPNSQSFIOQVOOTSSMV 287
Db 819 PKITGHQOQVTVQVVLKAGQPG-----AIIKTPV-----MSGVR 855
OY 288 IATCTTIVTSPVTVTVTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTVLH 347
Db 856 LVTVTVYSAVPAVTT-----LVKGT-----GVTTLG 884
OY 348 SVGPAAATGTTAGTGLQTSKPLVTSVA--NTVTVSLQPEKPVSGTAVTSLPAVTF 405
Db 885 TVTGTVST--SLAGAGHSTASLAPITTLTGATLSSQ---VINPATAVSAQTTL 938
OY 406 GETSGAALCLPSVAPVVSFCMDHICKPVTGTPQIKL-AQGPVLSOP 452
Db 939 TMAAGLTPTTITMQPV-----SQPQVTLTAPSGVEAP 973

RESULT 6
AMYL YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucosylase SI/52 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase) (1,4-alpha-D-glucan glucanhydrolase).
GN STAL OR STA2 OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;

```


CC OF SILKORM HEMOCYTIN.
 CC -1- SIMILARITY: CONTAINS 2 VMFC DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE (CTCK) DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
 CC or send an email to license@ebi.ac.uk).

DR EMBL: L21998; AAB95295.1; -;
 DR EMBL: M74027; AAB95875.1; -;
 DR EMBL: M94131; AAB95163.1; -;
 DR EMBL: M94132; AAB95164.1; -;
 DR Genew: HGNC:7512; MUC2.
 DR MIM: 158370; -;
 DR InterPro: IPR000359; Cys_knot.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002400; GF_cysknot.
 DR InterPro: IPR002919; TIL_cysrich.
 DR InterPro: IPR001007; VMF_C.
 DR InterPro: IPR001846; VMF_D.
 DR Pfam: PF00007; Cys_knot; 1.
 DR Pfam: PF00093; vwc; 1.
 DR Pfam: PF00094; vwd; 4.
 DR Pfam: PF01826; TIL; 1.
 DR PRINTS: PR00438; GRCYSKNOT.
 DR SMART: SM00214; vwc; 2.
 DR SMART: SM00216; vwd; 4.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS01185; CTCK_1; 1.
 DR PROSITE: PS01225; CTCK_2; 1.
 DR PROSITE: PS01208; VMFC; 2.
 KM Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 5179
 FT DOMAIN 1401 1747
 FT REPEAT 1401 1416
 FT REPEAT 1417 1432
 FT REPEAT 1433 1448
 FT REPEAT 1449 1464
 FT REPEAT 1465 1471
 FT REPEAT 1472 1478
 FT REPEAT 1479 1494
 FT REPEAT 1495 1517
 FT REPEAT 1518 1533
 FT REPEAT 1534 1536
 FT REPEAT 1537 1572
 FT REPEAT 1573 1596
 FT REPEAT 1597 1612
 FT REPEAT 1613 1635
 FT REPEAT 1636 1651
 FT REPEAT 1652 1675
 FT REPEAT 1676 1683
 FT REPEAT 1684 1699
 FT REPEAT 1700 1715
 FT REPEAT 1716 1731
 FT REPEAT 1732 1747
 FT REPEAT 1748 1763
 FT REPEAT 1764 1771
 FT REPEAT 1772 1787
 FT REPEAT 1788 1803
 FT REPEAT 1804 1819
 FT REPEAT 1820 1835
 FT REPEAT 1836 1851
 FT REPEAT 1852 1867
 FT REPEAT 1868 1883
 FT REPEAT 1884 1899
 FT REPEAT 1900 1915
 FT REPEAT 1916 1931
 FT REPEAT 1932 1947
 FT REPEAT 1948 1963
 FT REPEAT 1964 1979
 FT REPEAT 1980 1995
 FT REPEAT 1996 2011
 FT REPEAT 2012 2027
 FT REPEAT 2028 2043
 FT REPEAT 2044 2059
 FT REPEAT 2060 2075
 FT REPEAT 2076 2091
 FT REPEAT 2092 2107
 FT REPEAT 2108 2123
 FT REPEAT 2124 2139
 FT REPEAT 2140 2155
 FT REPEAT 2156 2171
 FT REPEAT 2172 2187
 FT REPEAT 2188 2203
 FT REPEAT 2204 2219
 FT REPEAT 2220 2235
 FT REPEAT 2236 2251
 FT REPEAT 2252 2267
 FT REPEAT 2268 2283
 FT REPEAT 2284 2299
 FT REPEAT 2300 2315
 FT REPEAT 2316 2331
 FT REPEAT 2332 2347
 FT REPEAT 2348 2363
 FT REPEAT 2364 2379
 FT REPEAT 2380 2395
 FT REPEAT 2396 2411
 FT REPEAT 2412 2427
 FT REPEAT 2428 2443
 FT REPEAT 2444 2459
 FT REPEAT 2460 2475
 FT REPEAT 2476 2491
 FT REPEAT 2492 2507
 FT REPEAT 2508 2523
 FT REPEAT 2524 2539
 FT REPEAT 2540 2555
 FT REPEAT 2556 2571
 FT REPEAT 2572 2587
 FT REPEAT 2588 2603
 FT REPEAT 2604 2619
 FT REPEAT 2620 2635
 FT REPEAT 2636 2651
 FT REPEAT 2652 2667
 FT REPEAT 2668 2683
 FT REPEAT 2684 2699
 FT REPEAT 2700 2715
 FT REPEAT 2716 2731
 FT REPEAT 2732 2747
 FT REPEAT 2748 2763
 FT REPEAT 2764 2779
 FT REPEAT 2780 2795
 FT REPEAT 2796 2811
 FT REPEAT 2812 2827
 FT REPEAT 2828 2843
 FT REPEAT 2844 2859
 FT REPEAT 2860 2875
 FT REPEAT 2876 2891
 FT REPEAT 2892 2907
 FT REPEAT 2908 2923
 FT REPEAT 2924 2939
 FT REPEAT 2940 2955
 FT REPEAT 2956 2971
 FT REPEAT 2972 2987
 FT REPEAT 2988 3003
 FT REPEAT 3004 3019
 FT REPEAT 3020 3035
 FT REPEAT 3036 3051
 FT REPEAT 3052 3067
 FT REPEAT 3068 3083
 FT REPEAT 3084 3099
 FT REPEAT 3100 3115
 FT REPEAT 3116 3131
 FT REPEAT 3132 3147
 FT REPEAT 3148 3163
 FT REPEAT 3164 3179
 FT REPEAT 3180 3195
 FT REPEAT 3196 3211
 FT REPEAT 3212 3227
 FT REPEAT 3228 3243
 FT REPEAT 3244 3259
 FT REPEAT 3260 3275
 FT REPEAT 3276 3291
 FT REPEAT 3292 3307
 FT REPEAT 3308 3323
 FT REPEAT 3324 3339
 FT REPEAT 3340 3355
 FT REPEAT 3356 3371
 FT REPEAT 3372 3387
 FT REPEAT 3388 3403
 FT REPEAT 3404 3419
 FT REPEAT 3420 3435
 FT REPEAT 3436 3451
 FT REPEAT 3452 3467
 FT REPEAT 3468 3483
 FT REPEAT 3484 3499
 FT REPEAT 3500 3515
 FT REPEAT 3516 3531
 FT REPEAT 3532 3547
 FT REPEAT 3548 3563
 FT REPEAT 3564 3579
 FT REPEAT 3580 3595
 FT REPEAT 3596 3611
 FT REPEAT 3612 3627
 FT REPEAT 3628 3643
 FT REPEAT 3644 3659
 FT REPEAT 3660 3675
 FT REPEAT 3676 3691
 FT REPEAT 3692 3707
 FT REPEAT 3708 3723
 FT REPEAT 3724 3739
 FT REPEAT 3740 3755
 FT REPEAT 3756 3771
 FT REPEAT 3772 3787
 FT REPEAT 3788 3803
 FT REPEAT 3804 3819
 FT REPEAT 3820 3835
 FT REPEAT 3836 3851
 FT REPEAT 3852 3867
 FT REPEAT 3868 3883
 FT REPEAT 3884 3899
 FT REPEAT 3900 3915
 FT REPEAT 3916 3931
 FT REPEAT 3932 3947
 FT REPEAT 3948 3963
 FT REPEAT 3964 3979
 FT REPEAT 3980 3995
 FT REPEAT 3996 4011
 FT REPEAT 4012 4027
 FT REPEAT 4028 4043
 FT REPEAT 4044 4059
 FT REPEAT 4060 4075
 FT REPEAT 4076 4091
 FT REPEAT 4092 4107
 FT REPEAT 4108 4123
 FT REPEAT 4124 4139
 FT REPEAT 4140 4155
 FT REPEAT 4156 4171
 FT REPEAT 4172 4187
 FT REPEAT 4188 4203
 FT REPEAT 4204 4219
 FT REPEAT 4220 4235
 FT REPEAT 4236 4251
 FT REPEAT 4252 4267
 FT REPEAT 4268 4283
 FT REPEAT 4284 4299
 FT REPEAT 4300 4315
 FT REPEAT 4316 4331
 FT REPEAT 4332 4347
 FT REPEAT 4348 4363
 FT REPEAT 4364 4379
 FT REPEAT 4380 4395
 FT REPEAT 4396 4411
 FT REPEAT 4412 4427
 FT REPEAT 4428 4443
 FT REPEAT 4444 4459
 FT REPEAT 4460 4475
 FT REPEAT 4476 4491
 FT REPEAT 4492 4507
 FT REPEAT 4508 4523
 FT REPEAT 4524 4539
 FT REPEAT 4540 4555
 FT REPEAT 4556 4571
 FT REPEAT 4572 4587
 FT REPEAT 4588 4603
 FT REPEAT 4604 4619
 FT REPEAT 4620 4635
 FT REPEAT 4636 4651
 FT REPEAT 4652 4667
 FT REPEAT 4668 4683
 FT REPEAT 4684 4699
 FT REPEAT 4700 4715
 FT REPEAT 4716 4731
 FT REPEAT 4732 4747
 FT REPEAT 4748 4763
 FT REPEAT 4764 4779
 FT REPEAT 4780 4795
 FT REPEAT 4796 4811
 FT REPEAT 4812 4827
 FT REPEAT 4828 4843
 FT REPEAT 4844 4859
 FT REPEAT 4860 4875
 FT REPEAT 4876 4891
 FT REPEAT 4892 4907
 FT REPEAT 4908 4923
 FT REPEAT 4924 4939
 FT REPEAT 4940 4955
 FT REPEAT 4956 4971
 FT REPEAT 4972 4987
 FT REPEAT 4988 5003
 FT REPEAT 5004 5019
 FT REPEAT 5020 5035
 FT REPEAT 5036 5051
 FT REPEAT 5052 5067
 FT REPEAT 5068 5083
 FT REPEAT 5084 5099
 FT REPEAT 5100 5115
 FT REPEAT 5116 5131
 FT REPEAT 5132 5147
 FT REPEAT 5148 5163
 FT REPEAT 5164 5179
 FT REPEAT 5180 5195
 FT REPEAT 5196 5211
 FT REPEAT 5212 5227
 FT REPEAT 5228 5243
 FT REPEAT 5244 5259
 FT REPEAT 5260 5275
 FT REPEAT 5276 5291
 FT REPEAT 5292 5307
 FT REPEAT 5308 5323
 FT REPEAT 5324 5339
 FT REPEAT 5340 5355
 FT REPEAT 5356 5371
 FT REPEAT 5372 5387
 FT REPEAT 5388 5403
 FT REPEAT 5404 5419
 FT REPEAT 5420 5435
 FT REPEAT 5436 5451
 FT REPEAT 5452 5467
 FT REPEAT 5468 5483
 FT REPEAT 5484 5499
 FT REPEAT 5500 5515
 FT REPEAT 5516 5531
 FT REPEAT 5532 5547
 FT REPEAT 5548 5563
 FT REPEAT 5564 5579
 FT REPEAT 5580 5595
 FT REPEAT 5596 5611
 FT REPEAT 5612 5627
 FT REPEAT 5628 5643
 FT REPEAT 5644 5659
 FT REPEAT 5660 5675
 FT REPEAT 5676 5691
 FT REPEAT 5692 5707
 FT REPEAT 5708 5723
 FT REPEAT 5724 5739
 FT REPEAT 5740 5755
 FT REPEAT 5756 5771
 FT REPEAT 5772 5787
 FT REPEAT 5788 5803
 FT REPEAT 5804 5819
 FT REPEAT 5820 5835
 FT REPEAT 5836 5851
 FT REPEAT 5852 5867
 FT REPEAT 5868 5883
 FT REPEAT 5884 5899
 FT REPEAT 5900 5915
 FT REPEAT 5916 5931
 FT REPEAT 5932 5947
 FT REPEAT 5948 5963
 FT REPEAT 5964 5979
 FT REPEAT 5980 5995
 FT REPEAT 5996 6011
 FT REPEAT 6012 6027
 FT REPEAT 6028 6043
 FT REPEAT 6044 6059
 FT REPEAT 6060 6075
 FT REPEAT 6076 6091
 FT REPEAT 6092 6107
 FT REPEAT 6108 6123
 FT REPEAT 6124 6139
 FT REPEAT 6140 6155
 FT REPEAT 6156 6171
 FT REPEAT 6172 6187
 FT REPEAT 6188 6203
 FT REPEAT 6204 6219
 FT REPEAT 6220 6235
 FT REPEAT 6236 6251
 FT REPEAT 6252 6267
 FT REPEAT 6268 6283
 FT REPEAT 6284 6299
 FT REPEAT 6300 6315
 FT REPEAT 6316 6331
 FT REPEAT 6332 6347
 FT REPEAT 6348 6363
 FT REPEAT 6364 6379
 FT REPEAT 6380 6395
 FT REPEAT 6396 6411
 FT REPEAT 6412 6427
 FT REPEAT 6428 6443
 FT REPEAT 6444 6459
 FT REPEAT 6460 6475
 FT REPEAT 6476 6491
 FT REPEAT 6492 6507
 FT REPEAT 6508 6523
 FT REPEAT 6524 6539
 FT REPEAT 6540 6555
 FT REPEAT 6556 6571
 FT REPEAT 6572 6587
 FT REPEAT 6588 6603
 FT REPEAT 6604 6619
 FT REPEAT 6620 6635
 FT REPEAT 6636 6651
 FT REPEAT 6652 6667
 FT REPEAT 6668 6683
 FT REPEAT 6684 6699
 FT REPEAT 6700 6715
 FT REPEAT 6716 6731
 FT REPEAT 6732 6747
 FT REPEAT 6748 6763
 FT REPEAT 6764 6779
 FT REPEAT 6780 6795
 FT REPEAT 6796 6811
 FT REPEAT 6812 6827
 FT REPEAT 6828 6843
 FT REPEAT 6844 6859
 FT REPEAT 6860 6875
 FT REPEAT 6876 6891
 FT REPEAT 6892 6907
 FT REPEAT 6908 6923
 FT REPEAT 6924 6939
 FT REPEAT 6940 6955
 FT REPEAT 6956 6971
 FT REPEAT 6972 6987
 FT REPEAT 6988 7003
 FT REPEAT 7004 7019
 FT REPEAT 7020 7035
 FT REPEAT 7036 7051
 FT REPEAT 7052 7067
 FT REPEAT 7068 7083
 FT REPEAT 7084 7099
 FT REPEAT 7100 7115
 FT REPEAT 7116 7131
 FT REPEAT 7132 7147
 FT REPEAT 7148 7163
 FT REPEAT 7164 7179
 FT REPEAT 7180 7195
 FT REPEAT 7196 7211
 FT REPEAT 7212 7227
 FT REPEAT 7228 7243
 FT REPEAT 7244 7259
 FT REPEAT 7260 7275
 FT REPEAT 7276 7291
 FT REPEAT 7292 7307
 FT REPEAT 7308 7323
 FT REPEAT 7324 7339
 FT REPEAT 7340 7355
 FT REPEAT 7356 7371
 FT REPEAT 7372 7387
 FT REPEAT 7388 7403
 FT REPEAT 7404 7419
 FT REPEAT 7420 7435
 FT REPEAT 7436 7451
 FT REPEAT 7452 7467
 FT REPEAT 7468 7483
 FT REPEAT 7484 7499
 FT REPEAT 7500 7515
 FT REPEAT 7516 7531
 FT REPEAT 7532 7547
 FT REPEAT 7548 7563
 FT REPEAT 7564 7579
 FT REPEAT 7580 7595
 FT REPEAT 7596 7611
 FT REPEAT 7612 7627
 FT REPEAT 7628 7643
 FT REPEAT 7644 7659
 FT REPEAT 7660 7675
 FT REPEAT 7676 7691
 FT REPEAT 7692 7707
 FT REPEAT 7708 7723
 FT REPEAT 7724 7739
 FT REPEAT 7740 7755
 FT REPEAT 7756 7771
 FT REPEAT 7772 7787
 FT REPEAT 7788 7803
 FT REPEAT 7804 7819
 FT REPEAT 7820 7835
 FT REPEAT 7836 7851
 FT REPEAT 7852 7867
 FT REPEAT 7868 7883
 FT REPEAT 7884 7899
 FT REPEAT 7900 7915
 FT REPEAT 7916 7931
 FT REPEAT 7932 7947
 FT REPEAT 7948 7963
 FT REPEAT 7964 7979
 FT REPEAT 7980 7995
 FT REPEAT 7996 8011
 FT REPEAT 8012 8027
 FT REPEAT 8028 8043
 FT REPEAT 8044 8059
 FT REPEAT 8060 8075
 FT REPEAT 8076 8091
 FT REPEAT 8092 8107
 FT REPEAT 8108 8123
 FT REPEAT 8124 8139
 FT REPEAT 8140 8155
 FT REPEAT 8156 8171
 FT REPEAT 8172 8187
 FT REPEAT 8188 8203
 FT REPEAT 8204 8219
 FT REPEAT 8220 8235
 FT REPEAT 8236 8251
 FT REPEAT 8252 8267
 FT REPEAT 8268 8283
 FT REPEAT 8284 8299
 FT REPEAT 8300 8315
 FT REPEAT 8316 8331
 FT REPEAT 8332 8347
 FT REPEAT 8348 8363
 FT REPEAT 8364 8379
 FT REPEAT 8380 8395
 FT REPEAT 8396 8411
 FT REPEAT 8412 8427
 FT REPEAT 8428 8443
 FT REPEAT 8444 8459
 FT REPEAT 8460 8475
 FT REPEAT 8476 8491
 FT REPEAT 8492 8507
 FT REPEAT 8508 8523
 FT REPEAT 8524 8539
 FT REPEAT 8540 8555
 FT REPEAT 8556 8571
 FT REPEAT 8572 8587
 FT REPEAT 8588 8603
 FT REPEAT 8604 8619
 FT REPEAT 8620 8635
 FT REPEAT 8636 8651
 FT REPEAT 8652 8667
 FT REPEAT 8668 8683
 FT REPEAT 8684 8699
 FT REPEAT 8700 8715
 FT REPEAT 8716 8731
 FT REPEAT 8732 8747
 FT REPEAT 8748 8763
 FT REPEAT 8764 8779
 FT REPEAT 8780 8795
 FT REPEAT 8796 8811
 FT REPEAT 8812 8827
 FT REPEAT 8828 8843
 FT REPEAT 8844 8859
 FT REPEAT 8860 8875
 FT REPEAT 8876 8891
 FT REPEAT 8892 8907
 FT REPEAT 8908 8923
 FT REPEAT 8924 8939
 FT REPEAT 8940 8955
 FT REPEAT 8956 8971
 FT REPEAT 8972 8987
 FT REPEAT 8988 9003
 FT REPEAT 9004 9019
 FT REPEAT 9020 9035
 FT REPEAT 9036 9051
 FT REPEAT 9052 9067
 FT REPEAT 9068 9083
 FT REPEAT 9084 9099
 FT REPEAT 9100 9115
 FT REPEAT 9116 9131
 FT REPEAT 9132 9147
 FT REPEAT 9148 9163
 FT REPEAT 9164 9179
 FT REPEAT 9180 9195
 FT REPEAT 9196 9211
 FT REPEAT 9212 9227
 FT REPEAT 9228 9243
 FT REPEAT 9244 9259
 FT REPEAT 9260 9275
 FT REPEAT 9276 9291
 FT REPEAT 9292 9307
 FT REPEAT 9308 9323
 FT REPEAT 9324 9339
 FT REPEAT 9340 9355
 FT REPEAT 9356 9371
 FT REPEAT 9372 9387
 FT REPEAT 9388 9403
 FT REPEAT 9404 9419
 FT REPEAT 9420 9435
 FT REPEAT 9436 9451
 FT REPEAT 9452 9467
 FT REPEAT 9468 9483
 FT REPEAT 9484 9499
 FT REPEAT 9500 9515
 FT REPEAT 9516 9531
 FT REPEAT 9532 9547
 FT REPEAT 9548 9563
 FT REPEAT 9564 9579
 FT REPEAT 9580 9595
 FT REPEAT 9596 9611
 FT REPEAT 9612 9627
 FT REPEAT 9628 9643
 FT REPEAT 9644 9659
 FT REPEAT 9660 9675
 FT REPEAT 9676 9691
 FT REPEAT 9692 9707
 FT REPEAT 9708 9723
 FT REPEAT 9724 9739
 FT REPEAT 9740 9755
 FT REPEAT 9756 9771
 FT REPEAT 9772 9787
 FT REPEAT 9788 9803
 FT REPEAT 9804 9819
 FT REPEAT 9820 9835
 FT REPEAT 9836 9851
 FT REPEAT 9852 9867
 FT REPEAT 9868 9883
 FT REPEAT 9884 9899
 FT REPEAT 9900 9915
 FT REPEAT 9916 9931
 FT REPEAT 9932 9947
 FT REPEAT 9948 9963
 FT REPEAT 9964 9979
 FT REPEAT 9980 9995
 FT REPEAT 9996 10011
 FT REPEAT 10012 10027
 FT REPEAT 10028 10043
 FT REPEAT 10044 10059
 FT REPEAT 10060 10075
 FT REPEAT 10076 10091
 FT REPEAT 10092 10107
 FT REPEAT 10108 10123
 FT REPEAT 10124 10139
 FT REPEAT 10140 10155
 FT REPEAT 10156 10171
 FT REPEAT 10172 10187
 FT REPEAT 10188 10203
 FT REPEAT 10204 10219
 FT REPEAT 10220 10235
 FT REPEAT 10236 10251
 FT REPEAT 10252 10267
 FT REPEAT 10268 10283
 FT REPEAT 10284 10299
 FT REPEAT 10300 10315
 FT REPEAT 10316 10331
 FT REPEAT 10332 10347
 FT REPEAT 10348 10363
 FT REPEAT 10364 10379
 FT REPEAT 10380 10395
 FT REPEAT 10396 10411
 FT REPEAT 10412 10427
 FT REPEAT 10428 10443
 FT REPEAT 10444 10459
 FT REPEAT 10460 10475
 FT REPEAT 10476 10491
 FT REPEAT 10492 10507
 FT REPEAT 10508 10523
 FT REPEAT 10524 10539
 FT REPEAT 10540 10555
 FT REPEAT 10556 10571
 FT REPEAT 10572 10587
 FT REPEAT 10588 10603
 FT REPEAT 10604 10619
 FT REPEAT 10620 10635
 FT REPEAT 10636 10651
 FT REPEAT 10652 10667
 FT REPEAT 10668 10683
 FT REPEAT 10684 10699
 FT REPEAT 10700 10715
 FT REPEAT 10716 10731
 FT REPEAT 10732 10747
 FT REPEAT 10748 10763
 FT REPEAT 10764 10779
 FT REPEAT 10780 10795
 FT REPEAT 10796 10811
 FT REPEAT 10812 10827
 FT REPEAT 10828 10843
 FT REPEAT 10844 10859
 FT REPEAT 10860 10875
 FT REPEAT 10876 10891
 FT REPEAT 10892 10907
 FT REPEAT 10908 10923
 FT REPEAT 10924 10939
 FT REPEAT 10940 10955
 FT REPEAT 10956 10971
 FT REPEAT 10972 10987
 FT REPEAT 10988 11003
 FT REPEAT 11004 11019
 FT REPEAT 11020 11035
 FT REPEAT 11036 11051
 FT REPEAT 11052 11067
 FT REPEAT 11068 11083
 FT REPEAT 11084 11099
 FT REPEAT 11100 11115
 FT REPEAT 11116 11131
 FT REPEAT 11132 11147
 FT REPEAT 11148 11163
 FT REPEAT 11164 11179
 FT REPEAT 11180 11195
 FT REPEAT 11196 11211
 FT REPEAT 11212 11227
 FT REPEAT 11228 11243
 FT REPEAT 11244 11259
 FT REPEAT 11260 11275
 FT REPEAT 11276 11291
 FT REPEAT 11292 11307
 FT REPEAT 11308 11323
 FT REPEAT 11324 11339
 FT REPEAT 11340 11355
 FT REPEAT 11356 11371
 FT REPEAT 11372 11387
 FT REPEAT 11388 11403
 FT REPEAT 11404 11419
 FT REPEAT 11420 11435
 FT REPEAT 11436 11451
 FT REPEAT 11452 11467
 FT REPEAT 11468 11483
 FT REPEAT 11484 11499
 FT REPEAT 11500 11515
 FT REPEAT 11516 11531
 FT REPEAT 11532 11547
 FT REPEAT 11548 11563
 FT REPEAT 11564 11579
 FT REPEAT 11580 11595
 FT REPEAT 11596 11611
 FT REPEAT 11612 11627
 FT REPEAT 11628 11643
 FT REPEAT 11644 11659
 FT REPEAT 11660 11675
 FT REPEAT 11676 11691
 FT REPEAT 11692 11707
 FT REPEAT 11708 11723
 FT REPEAT 11724 11739
 FT REPEAT 11740 11755
 FT REPEAT 11756 11771
 FT REPEAT 11772 11787
 FT REPEAT 11788 11803
 FT REPEAT 11804 11819
 FT REPEAT 11820 11835
 FT REPEAT 11836 11851
 FT REPEAT 11852 11867
 FT REPEAT 11868 11883
 FT REPEAT 11884 11899
 FT REPEAT 1190


```

DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Calphotin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Canlon-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Canlon-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
OF CA+2 PER MOL OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
COMPOUND EYES AND OCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
DEVELOPMENT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL: L02111; AAA28405.1; -
DR EMBL: L05080; AAA28420.1; -
DR PIR: A47282; A47282.
DR Flybase: FBgn0010218; Cpn.
KW Calcium-binding.
FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 T -> V (IN REF. 2).
FT CONFLICT 76 76 T -> A (IN REF. 2).
FT CONFLICT 100 100 P -> PP (IN REF. 2).
FT CONFLICT 126 127 VO -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 160 160 S -> E (IN REF. 2).
FT CONFLICT 534 534 I -> T (IN REF. 2).
FT CONFLICT 689 689 T -> L (IN REF. 2).
FT CONFLICT 703 703 D -> E (IN REF. 2).
FT CONFLICT 721 721
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CEE CRC64;

Query Match 8.6%; Score 194; DB 1; Length 865;
Best Local Similarity 22.2%; Pred. No. 0.00045;
Matches 121; Conservative 71; Mismatches 209; Indels 144; Gaps 23;

```

```

QY 105 KYAAVTPVKKL-AQIGTIVTVTPKSSVQSVAVPVSVT-VTPGKPLNTVTLKPSLQ 162
DB 116 PVAAAPPTPVQIPVAVPVLATPPVAAADPTPAVPPVISPVIASPPVPANPTVPAAP 175
QY 163 ASSTPSNEPMILKAENSAVQINLSPTMLE-----NVKCKNPLAMLIKACS 209
DB 176 VAAVPAVPPVAVPVLAAVAPAAVPAVPAVEPPAPPAEIVAVIPEC---VAPLI----- 227
QY 210 GSQSPKMGQNVKRLVEQLDAKIEAEETFRKLYVELKSSDPHLVPELKKSVALARQLP 269
DB 228 ----PEVSVATKRPILAAAEPPVVAAPATETPPVAPAAASPHVSAVPAVETAAVVA-----P 278
QY 270 NSQSFIOOCVOQTSSSDKVIATCTTTTTSPTVTTSSQSEKSIIV-----SGATAPRV 325
DB 279 VSAS-----TEPPVAAATLTTPAPT-PALAPVVAESQVAAANTVATPTTPAPEPTI 329
QY 326 S-----VQTLNPLAGPVGAKAGVTVTHSVGTATGTTAGTGLQTSKP 370
DB 330 APPVVAETPEVSAVVAETPPVPPVVAAS-----IPAPVAT-----TPVP 372
QY 371 LVTSVANTVTTVTSLOPEKP-----VVSQTAATLS-----LPATTEGET- 408
DB 373 ATLAVTDPDTASAVPELPVIAAPVSAVPAETPVDLAPVLPVVAEEVPAVVAETP 432
QY 409 -----SGAALCPSPKPPVVSFCWD-----HICKVYIGTP-----VQIKLAQCP 447
DB 433 ETPAPASAPVTIALDLIPEVAPVIAAPSDAPAPASAAADISTPPTTASVETTPAPAA 492
QY 448 VLSQP 452
DB 493 VPTEP 497

RESULT 9
VG50_HEV11
ID VG50_HEV11 STANDARD; PRT; 670 AA.
AC 000130;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical gene 50 protein.
GN 50.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Ictalurid Herpes-like viruses.
OX NCBI_TaxID=10401;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Aburn 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL: M75136; AAA88153.1; -
DR PIR: F36791; F36791.
KW Hypothetical protein; Repeat.
FT REPEAT 143 158 1.
FT REPEAT 171 186 2.
FT REPEAT 200 214 3.
FT REPEAT 215 233 4.
FT REPEAT 234 252 5.
FT REPEAT 253 268 6.
FT REPEAT 279 293 7.

```


DE Hypothetical protein ZK945.9 in chromosome II.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Wilkinson-Sprat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z48544; CAB70192.1; -
DR EMBL; Z48582; CAB70192.1; JOINED.

DR	EML1: Z48544; CAB70201.1; JOINED.
DR	Wormpep: ZK945.9; CE25697.
DR	InterPro: IPR003111; cat_channel_TrypL.
DR	InterPro: IPR001024; Lipoxigenase_LH2.
DR	InterPro: IPR000636; M-channel_nlg.
DR	InterPro: IPR000203; PKD_cys_rich.
DR	Pfam: PF00520; Ion_trans. 1.
DR	Pfam: PF01477; PLAT. 1.
DR	Pfam: PF01825; GPS. 1.
DR	SMART: SM00303; GPS. 1.
DR	SMART: SM00308; LH2. 1.
KW	Hypothetical protein; Transmembrane.
FT	DOMAIN 266 1196
FT	DOMAIN 1105 1241
FT	DOMAIN 2071 2120
FT	TRANSMEM 13 30
FT	TRANSMEM 51 73
FT	TRANSMEM 2139 2161
FT	TRANSMEM 2348 2367
FT	TRANSMEM 2390 2412
FT	TRANSMEM 2451 2468
FT	TRANSMEM 2483 2505
FT	TRANSMEM 2567 2589
FT	TRANSMEM 2836 2858
FT	TRANSMEM 2939 2961
FT	TRANSMEM 2976 2998
FT	TRANSMEM 3038 3060
SO	SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;
Query Match	
Best Local Similarity 7.88; Score 175.5; DB 1; Length 3178;	
Matches 100; Conservative 62; Pred. No. 0.025; Mismatches 215; Indels 63; Gaps 16;	
QY	2 TLVNVKVAVSAAPRVSSSPRLPAPQIVAKPANTTTIOFPANLQLPQGVTLIKNSGLM 61
DB	356 TLVLSIPPTTTPETSTLSSLPDNAICSYLDETTSTFTTMTLTSTTEPSPSTTTE 415
QY	62 LVSPQQTTRAEFTSNITSPAPVA--NPQTVCICTVPNSSSLIKAKVATPVKKLAIG 119
DB	416 VTSVSSVTTTEPTTLTSTASTSTETPSTSVTTPSTST-----PVSTVSS 465
QY	120 TTVVTVTPKPSVOSVAN-PTSVTVTVTGKRLNVYTLKPSLGLASTSPNEPIKAENS 178
DB	466 SSSSTVTVTPSTSTSPSPSVTSTTAP--STSTGPPS--SSSTPSSSTASSSVST 521

[illegible]

	RESULT	12		
CC	VGLX_HSVEB			
ID	VGLX_HSVEB	STANDARD:	PRT:	797 AA.
AC	p28968;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	01-DEC-1992 (Rel. 24, Last annotation update)			
DE	Glycoprotein x precursor.			
CN	71			
OS	Equine herpesvirus type 1 (strain Ab4p) (EHV-1);			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Alphaherpesvirinae; Varicellovirus.			
OX	NCBI_TaxID=31520;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92295566; PubMed=1318606;			
RA	Telford E.A.R., Watson M.S., McBride K., Davison A.J.;			
RT	"The DNA sequence of equine herpesvirus-1.";			
RL	Virology 189:304-316(1992).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and that statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			

Db	134	TAVTTAASTSEFTTATATATSTPTTTTPTSTTTTATTTTPTTATTTTAA	193
Qy	181	VOINLSPMLNEVAKKCNFLAMLKLKLCSSQSQSEMGQNKVLVEQLDCKIAEETR	240
Db	194	TTTATTTTAAATTTA-----ATTAAATTTAA	218
Qy	241	LYVELKSSPOPHLVPFLKKSVALROLPLNSQSFIOCCVOQTSNDVIATCTTTVTS	300
Db	219	TTTAAATTS-----SATAATTAATTAATTAATTAATTAATTTA	256
Qy	301	VTTTSSSSOSEKSIIVSGATAPRTVSQVTLINPLAGPVGARAGVTLHSGVPTAATG	360
Db	257	ATTGSPGSGSTPTTGASTSPSSASTATSTSTSTSAATST-----PTPTSAT	311
Qy	361	-GTGLQTSKRLV--TSVANTVYTSLODEKPKVSGAVTSLSLAVYT	404
Db	312	ESTTEAPSTPTTDTPTTPTSEATTATNTSPESVTYSASTSTSTTTAAFT	357

RESULT 13			
ID	VEP3_EBVAB	STANDARD;	PRT; 886 AA.
AC	007284;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	Envelope glycoprotein GP340 (Membrane antigen) (MA).		
GN	BLRF1.		
OS	Epstein-barr virus (strain AC676) (human herpesvirus 4), and		
OS	Epstein-barr virus (strain P3HR-1) (human herpesvirus 4).		
CC	Viruses; dsDNA viruses; no RNA stage; Herpesviridae;		
CC	Gammaherpesvirinae; Lymphocryptovirus.		
OX	NCBI_TaxID=82830, 82829;		
LN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-AC676, and P3HR-1;		
RX	MEDLINE=93331716; PubMed=8393237;		
RA	Lees J.F., Arrand J.E., Pepper S.V., Stewart J.P., Mackett M.,		
RT	Arrand J.R.;		
RT	"The Epstein-Barr virus candidate vaccine antigen gp340/220 is highly		
RT	conserved between virus types A and B."		
RL	Virology 195:578-586 (1993).		
LN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-P3HR-1;		
RL	Klein K., Mueller-Lantzsch N.;		
RL	Submitted (OCT-1992) to the EMBL/GenBank/DDAJ databases.		
CC	-1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN		
CC	B-CELLS.		
CC	-1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL		
CC	ENVELOPE.		

```

FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 533 533 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 589 589 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 680 680 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 886 AA; 92388 MW; 4394F6130DEC8A CRC64;

```

Query Match 7.6%; Score 172; DB 1; Length 886;

Best Local Similarity 23.4%; Pred. No. 0.0084; Mismatches 194; Indels 118; Gaps 22;

```

Matches 111; Conservative 51; Mismatches 194; Indels 118; Gaps 22;

OY 2 TLVTKYAPVAPPKVSSGRLPAPQIVAVKAPNTT-----IQFANLQLPFGVLKSN 56
DB 414 TTHHKVIFSKAPRSTSTSTL---NTTGRAAPRTTGLBSTHVPINLTPAST----- 464
OY 57 SGPLMLVSPQOIVTRAETTSNITSRAVPA-----NPOTVICTVFNSS 100
DB 465 -GPTVSTADVTSPAGTTS--ASPVTSPSPRDNGTESKADMDSPSVAVTPPNNAT 521
OY 101 SOLIKRVAVTPVKKLAOIGTVVY---TVPKP---SSVGVANPTSVVY--TGK--PLN 151
DB 522 SPTPAVATTPPNATSPTLGKTSPTSAVTTPPNATSPTPPNATPTLGKTSPTS 581
OY 152 TVTTLKPS---SLGASSTPSNEPNLKAENSAVINLSP-----TLENVKCKN 198
DB 582 AVTTPPNATSPVGETSTPQANTNTNHTLGSTSPVTVSPKNATSAVTTGONNITSSST 641
OY 199 FLAMLIKACSGSPQSPKQVAKVLEQDLDAKIEAEETFRKLYVELKSSPOPHLVEFK 258
DB 642 SSMKLRPSSISLETLSPTSDNSTS-----HMLPLTSAHP----- 675
OY 259 KSVVALROLPLNSQSFIOOCVOOTSSDMVIACTTIVTTSPPVTTYS---SSQSK-- 312
DB 676 -----TGGENTIOVTPASTSTHHV---STSSPAPRPGTTSQASGNSSTSTKPG 722
OY 313 SIIVSGATAPRTVSQVTLNPLAGVAKAGVTLHSGVPTA--ATGGT--TAGTGLQTSK 369
DB 723 EVNVTKGTGTPPKNT-----SPQA--PSCOKTAVPVITSTGCKANSTTGCKHTTGCAKSTE 777
OY 370 PLVTSVANTVT-----TVSLOPEKPVVSGVAVTSLSLPAVTFGETSGAICLP 416
DB 778 PTTDXGDSSTTRTRYNATVYLPSTSSKLPRMFTSPVY---TAQATVPVP 828

RESULT 14
YAG3-YEAST STANDARD; PRT; 1322 AA.
AC P39712;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 138.1 KDa protein in FLO9-GDH3 Intergenic precursor.
GN YAL063C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
MEDLIN=95249563; PubMed=7731988;

```

```

RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -1- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U12980; AAC04971.1; -
DR SGD: S0000059; YAL063C.
DR InterPro: IPR001389; Flocculin.
DR Pfam: PF00624; Flocculin; 13.
KW Hypothetical protein; Glycoprotein; Transmembrane; signal.
FT CHAIN 1 24
FT SIGNAL 1 24
FT TRANSMEM 366 388
FT TRANSMEM 754 775
FT CARBOHYD 135 135
FT CARBOHYD 187 187
FT CARBOHYD 203 203
FT CARBOHYD 257 257
FT CARBOHYD 262 262
FT CARBOHYD 270 270
FT CARBOHYD 329 329
FT CARBOHYD 419 419
FT CARBOHYD 464 464
FT CARBOHYD 509 509
FT CARBOHYD 554 554
FT CARBOHYD 599 599
FT CARBOHYD 644 644
FT CARBOHYD 689 689
FT CARBOHYD 734 734
FT CARBOHYD 888 888
SQ SEQUENCE 1322 AA; 138072 MW; AADFDLFF13267CEA CRC64;

```

```

Query Match 7.6%; Score 172; DB 1; Length 1322;
Best Local Similarity 21.0%; Pred. No. 0.014; Mismatches 164; Indels 128; Gaps 21;
Matches 97; Conservative 73; Mismatches 164; Indels 128; Gaps 21;

OY 22 LPAPO-IVAVKADNTTIOF-----PANLQLPFGVLKSNSG-----PLMLVSPQOT 68
DB 663 LPIDETIIVIRTPYATATTATTQPPNDTSTSTETITVGTNGLPDFTIIVIRPTT 722
OY 69 VTRAETT-----SNITSRAVPANPOTVKTICTVFNSSQDLKRAVATPVK 113
DB 723 ATTAATTTQPMNDTSTSTETITVGTNGVPIVD-ETVIAIIRPTSE----- 768
OY 114 KLAOIGTVVTPPKBSVSVAVPVTVTGKPLN-IVTTLK-PSSLGASTTSNEP 171
DB 769 -----GLISTTEPMWGTFTSTETMTVGTNGQPDFTVIAIRPTSGLVTT-TTEP 822
OY 172 -----NLKAENSAVAOINSPMLENVKCKNFMALIK---LACSGSPQPMGVNKK 222
DB 823 WTGTFTSTETMTTIGTNQPI-----DEVIIVATPTTAISSSSSSGQ----- 869
OY 223 LVQLDLDAKIEAEETFRKLYVELKSSPOPHLVEFKSVVALROLPLNSQSFIOOCVQOT 282
DB 870 -----ITSFTSARPIITPFYPS-----NGSVISSSVISS 900
OY 283 S--SDMVIACTTT--VTTSPVTTTVSSQSKSIIVSGATAPRVSVQTLNPLAGPVG 338
DB 901 SDTSSLVISSSVTSSTVSSPVISSSFSSP-----VISSTTSASITLSES----- 946
OY 339 AKAGVATLHVGPTAATGCTAGTGLQTSKPLVTSVA--NTVTVSLOPEKPVVSGTAV 396

```

Db 947 SKSVIPIRSS--STGSGSESTGASASSSSSTISSEPKSTYSSSLPVTSATTSOI 1004

Oy 397 TSLPAAVTEGTGAACICLPSPKPVSEFCMDHICKPVIGPV 438

Db 1005 TSSLPVTTTSTSEOTLV-----TWSCHSECHTESISSAI 1041

RESULT 15

ZAN_HUMAN STANDARD; PRT; 2700 AA.

AC Q9Y493; 000218; (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Zonadhesin (Fragment).

GN ZAN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE OF 1-2379 FROM N.A.

RA MEDLINE=99018118; PubMed=9799793;

RA Glockner G., Scherer S., Schattteoy R., Boright A., Weber J.,

RA Tsui L.C., Rosenthal A.;

RT "Large-scale sequencing of two regions in human chromosome 7q22:

RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci

RT reveals 17 genes.";

RL Genome Res. 8:1060-1073(1998).

RN [2]

RP SEQUENCE OF 2338-2700 FROM N.A.

RC TISSUE=Testis;

RA MEDLINE=97271566; PubMed=9126492;

RA Gao Z., Harumi T., Garbers D.L.;

RT "Chromosome localization of the mouse zonadhesin gene and the human

RT zonadhesin gene (zan).";

RL Genomics 41:119-122(1997).

CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA

CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR

CC SIGNALING.

CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE

CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.

CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE

CC ZONA PELLUCIDA.

CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,

CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF

CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.

CC -1- DOMAIN: THE WMPD DOMAINS 2 AND 3 MAY MEDIATE COVALENT

CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).

CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.

CC -1- SIMILARITY: CONTAINS 4.5 WMPD DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF053356; AAC78790.1; -

DR EMBL; U83191; AAC51208.1; -

DR GENB; HGNC:12857; ZAN.

DR MIM; 602372; -

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR000998; MAM_domain.

DR InterPro; IPR002919; TIL_Cysrich.

DR InterPro; IPR003328; TILA_Cysrich.

DR InterPro; IPR001846; WMP_D.

DR Pfam; PF00094; wcd; 4.

DR Pfam; PF00629; MAM; 4.

DR Pfam; PF01826; TIL; 5.

DR Pfam; PF02345; TILA; 4.

DR SMART; SM00216; WMD; 1.

DR PROSITE; PS01186; EGF_2; 3.

DR PROSITE; PS00740; MAM_1; 1.

DR PROSITE; PS50060; MAM_2; 4.

DR GlycoProtein; Transmembrane; Cell adhesion; Repeat.

KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.

FT NON_TER 1 109

FT DOMAIN <1 109

FT DOMAIN 112 136

FT DOMAIN 161 326

FT DOMAIN 322 446

FT DOMAIN 483 951

FT DOMAIN 953 1065

FT DOMAIN 1066 1454

FT DOMAIN 1455 1861

FT DOMAIN 1862 2292

FT DOMAIN 2293 2684

FT DOMAIN ? ?

FT CARBOHYD 74 74

FT CARBOHYD 403 403

FT CARBOHYD 1023 1023

FT CARBOHYD 1099 1099

FT CARBOHYD 1618 1618

FT CARBOHYD 1737 1737

FT CARBOHYD 1832 1832

FT CARBOHYD 1878 1878

FT CARBOHYD 2136 2136

FT CARBOHYD 2505 2505

FT CARBOHYD 2374 2379

FT NON_TER 2700

SO SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;

Query Match 7.68; Score 171.5; DB 1; Length 2700;

Best Local Similarity 21.08; Pred. No. 0.034;

Matches 104; Conservative 60; Mismatches 178; Indels 153; Gaps 22;

Oy 9 PVSAAPPKVSQBPRLPA--POLIVAKAPNTTITOPANQLPGRVYLINSGLPLMLSPQ 66

Db 518 PSEKPNMSEKPTIPSEKPTITLTERP-----TIPSEKPTIPSEKPTISTEKPTVPEEP 571

Oy 67 QTVTRAET-----SNITSPAVPANOQTVK-----IC----- 94

Db 572 TTPEETTTMEPEVPIPEKPSIPEKPSIPEKPTIMEETISTEPTICPEKPTIPT 631

Oy 95 ---TVPNSSQLIKKAVTPVK--LADIGTTVVT---TVP--KPS-SVQSAVAPTSTVT 143

Db 632 EKPTIPTEKSTIPSEKPTTPEKPTIPTEKPTIPTEKPTIPTEKPTIPTEKLT 691

Oy 144 VTGPKPLNTVTTLPKS-SLGASSTPSPNEPNLKAEN-SAAVOINSPMLNENKCKKNFLA 201

Db 692 IPTEKP-TIPTEKPTISTEETPTTEETPTISTEKPSIPMEKPTLPPT----- 736

Oy 202 MLIKIACSGSPQENQVKKIVOLDPAKIEAEFTKLVLELKS---SPQPHLVPLK 258

Db 737 -----BETTSVSEETISTEKLTIPEKPTISTEKPTIPTEK 773

Oy 259 KSVVALROLPLNSQSFIOQCVOQTSSDNVIAFTCTTVTSPVTTVSSSQ---SEKSI 314

Db 774 PTISPEKLTIPTEK-----LTIPTEKPTIPIETTTISTEKLTIPTTEKPT 817

Oy 315 IVSGATAPRTVSQTLNPLNLPAGVGAAGVTLHSGVPAAGGTAGGLG--QTSKRLV 372

Db 818 I-----SEKPTISTEKP-----TIPTEKPTIPEETTTISTEKLTIPEKPTI 860

Oy 373 TSVANTVTVTSIAPPEKPVVSGTAVTSLPAVTEGTGAACICLPSPKPVSEFCMDHICKP 432

Db 861 SPEKLTIP-----EKPTISTEKPTIPTEKLT-----IPTEKPTIP-----TEKP 900

Oy 433 VIGTPVOIKLQAPGP 447

Db 901 TIPEKLTALRPAP 915

Thu Feb 20 16:36:28 2003

us-09-763-909-2_copy_1_452.rsp

Page 14

Search completed: February 16, 2003, 21:56:12
Job time : 19.7109 secs

C:Genetics:
A:Gene: MUC5B

Query Match 8.2%; Score 185.5; DB 2; Length 3570;
Best Local Similarity 22.0%; Pred. No. 0.0081;
Matches 100; Conservative 53; Mismatches 204; Indels 97; Gaps 14;

```

OY 2 TLTVKAVAPVAPKSSGRLPAPOLVAVKAPNTTIOFPANLQLPFGTVLKNSGPILM 61
DB 3179 TLVLTATATGATGSAVAFSSSTPGTAHTKVPPTTTGTATPSSPGTAL-----TPPW 3234
OY 62 LVSPOQTVAETTSNITSRP-AVPAAPOTVAKICTVPNSSSQLIKKVAATPVKKLAQIGT 120
DB 3235 ISTTTPTPTTPTTSGSTVPSSIPGTHT-----ARVLT 3269
OY 121 TVTTPKPSVSVAVPTSVTVTPGKPLNTVTTLKPSLSGASTPSNEPNLKAENSA 180
DB 3270 TTTTIV-----ATGSMATPSSSTQTSCTPPLTTTATATATGSGTTPSGTPTTIPPLV 3325
OY 181 VOINISPTMLENVKCKKNFLAMLIKIACSGSQSPENGVKKLVEQLDAKIEAEFTTK 240
DB 3326 TSMATTPAATSS-----KATSSSPRTATLPLVLTSTA--TKSTATST-- 3367
OY 241 LVELKSSQPHLVPELKRKSVVALRQLPNSQSFIOQCYOOTSDDAVIATCT--TVTTS 298
DB 3368 -----PIPSSTLTMTWTPA-----QTTPMSTMTSTHTSTPTTHTS 3406
OY 299 PVTTTSSSSOSEKSIIVSGATAPRTVSQTLNPLAGPVAKAGVYTLHS----- 348
DB 3407 TVLTATATATGATGSAVAFSSSTPGTAHTKVPPTTTGTATPSSPGTAL-----TPPW 3466
OY 349 -----VGPATAGGTAGTGLQTSKPLVTSVANTVT-TVSLQPEKPVSGTAVTSL 400
DB 3467 PSTIATVAVPTGTAATSSSTLGTATHTPK-VYAMATMPATATASVSSSTVGTTRPAVL 3525
OY 401 PA-VTFGFTSGAICLPKPVKPVVSCMDHICKP 432
DB 3526 PSSLPFTSVSVSSSVLTTLRP-TGFPSSHFTSP 3558

```

RESULT 15

PC4395

mucin 3 - human (fragment)
C:Species: Homo sapiens (man)

C:Date: 10-Nov-1997 #sequence_revision 10-Nov-1997 #text_change 08-Oct-1999
C:Accession: PC4395

R:Van Klinken, B.J.W.; Van Dijk, T.C.; Ounsoren, E.; Buelter, H.A.; Dekker, J.; Elnert

Biochem. Biophys. Res. Commun. 238, 143-148, 1997

A:Title: Molecular cloning of human MUC3 cDNA reveals a novel 59 amino acid tandem repeat

A:Reference number: PC4395; MUID:97445141; PMID:9299468

A:Accession: PC4395

A:Molecule type: mRNA

A:Residues: 1-648 <VAN>

A:Cross-references: DBJ:AF016692; NID:g2454614; PIDN:ABJ71685.1; PID:g2454615

A:Experimental source: Intestine

F:1-59,60-118,119-177,178-236,237-295/Region: repeat

F:296-565/Region: semi-unique #status predicted

F:566-582,583-600,616,617-633,634-647/Region: repeat

Query Match 8.2%; Score 185; DB 2; Length 648;

Best Local Similarity 22.8%; Pred. No. 0.0092;

Matches 102; Conservative 63; Mismatches 188; Indels 94; Gaps 18;

```

OY 4 VTKVAPVAPVSSGRLPAPOLVAVKAPNTTIOFPANLQLPFGTVLKNSGPILMV 63
DB 137 VDRSTPTVTSQSNSTP--TPREVITLPMSTPSEVSTPLT--MPVSTTVTTSAGTAST 193
OY 64 SPOQTVAETTSNITSRPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 123
DB 194 LVDVSTVPTVTSQSNSTP--TPREVITLPMSTPSEVSTPLT--MPVSTTVTTSAGTAST 242
OY 124 T-----TVKPSVSVSVAVPTSVTVTPGKPLNTVT-----TLKPSLSGASTPSNEPNL 173

```

```

DB 243 TSSEGSTLSTPSPVSTPVTSTEAISSATLSDTSMVSMPEISTLGTTLVSTTPV 302
OY 174 KAENSAVOINISPTMLENVKCKKNFLAMLIKIACSGSQSPENGVKKLVEQLDAKIE 233
DB 303 RPEESTPST--PSV-----YTSMTTASGSSSP-----T 332
OY 234 ABEFTKRLVELKSSQPHLVPELKRKSVVALRQLPNSQSFIOQCYOOTSDDAVIATCT 293
DB 333 TLEGTMTMSTSTSEKSTLITVL--ISPISVSPSEASTLSTPPGDDSTPLTSTKAG 389
OY 294 TVTTSPTVTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVAKAGVYTLHSVGP 353
DB 390 SFSIPAEVTTIRISITSEK-----TPLTTLVST-----TL---PTS 424
OY 354 ATGTTAGTGLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLPVTGFTSGAAI 413
DB 425 FPGASIASPTPLDTSTTFPS-----TDASTPTIV-----ATTISVYTBESTGT 475
OY 414 CLPSKPVVSCMD--HICKPVIGTPV 438
DB 476 FIPST-PTVSTADVPFATGAVSTPV 501

```

Search completed: February 16, 2003, 22:00:40
Job time : 25.2925 secs

```
Db      517 TAKPGANGCTVTTTTAKPTGANGCTVTTTAAK--TGATGTGVTATTAAKPTG- -ANGFVWT- 571
          : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      399 SUPAVTFEGTSGA--AICLPSVKP 420
          | : | | | : | |
Db      572 ----TTAKPAGANGCTVTTTTAAK 590

RESULT 12
T29634
hypothetical protein C12D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T29634
R:Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid C12D12.
A:Reference number: Z20656
A:Accession: T29634
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1825 <NHA>
A:Cross-references: EMBL:U51998; PIDN:AAA96080.1; GSPDB:GN00028; CESP:C12D12.1
A:Experimental source: strain Bristol N2; clone C12D12
C:Genetics:
A:Gene: CESP:C12D12.1
A:Map position: X
A:Insertions: 48/1; 86/3; 137/1, 172/3; 224/3; 253/1; 287/3; 328/2; 454/1; 487/3; 692/2/1
C:Superfamily: Epstein-Barr virus membrane antigen gp350

Query Match           8.5% Score 190.5; DB 2; Length 825;
Best Local Similarity 21.8%; Pred. NO. 0.00059;
Matches   98; Conservative    44; Mismatches 169; Indels 139; Gaps 16;

Qy      4 VTKVAPVSAPKVSSCGRLPAPOIYAVKAPNTTITQIPAN-IDLPQTVLIKNSGLML 62
          |||::||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      472 VTTVSFM-S-PRTVVIVPTTPRP-----VPTTNFP-PANPTATPTTV---GTSKOTNF 520
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      63 VSQP-QYTVAETSNITSIRPVPNPQVIKICTV--PNSSQLIKKVAVTPEKKIAQI 118
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      521 ISPHSTIGSIIVSTPRMAPOTSASPTTTPHTTAQSOPTTKPVVYVTVNSVTP----- 573
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      119 GTTVTVTVPKPSVSVSAVPISGVVTPGPKPLMTVTTLKPSLSGASTPSENEINAKENS 178
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      574 -STGTTYVFPVPTTGSPTTQITAPVTKPPVPSSTGTAPP----VTPISQPPVT--- 623
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      179 AAVQINSPTMLENVKKCNFLAMLIKLAGSCSQSDPMGMVKLEQLIDAKEEFT 238
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      624 -----T 624

Qy      239 RKLYELKSSPOPHLVPLELKSSVALROLPLNSOSFIQQCVQOISSDMVIACTTT--- 294
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      625 TSLLTTLTTPYP-----VTTVPPSSAT-----VPTTPPYTVVAATTSKAP 668
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      295 -VTTSPV-V-TTVSSOSEKSIIYGATAPRVSYQILNLPLAGPVGAKGVTILMSVCP 351
          |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      669 VTTSTPLAPIPTKLPISTPSTVGTSPTAPANLTPTTAPVANPTSSTAAPTAVNPSTP 728
          |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      352 TAATGGTAGTGLDTSKPLVTSVANTVTVSSLQEPKEPV----- 392
          |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      729 TTAPVPVPVPTTTPTTTTSTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 788
          |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      393 ---GTAATLSLPAVFEGTSGAAILGPSV 418
          |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      789 ATGPATPTPTTIPTTT-GSSNAVVALPV 817
          |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 13
F96615
Probable Myb-family transcription factor F16M22.4 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: F96615
```


[illegible]

```

RESULT 10
T34369
hypothetical protein T19D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34369
R:Favella, A.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T19D12.
A:Reference number: 221513
A:Accession: T34369
A:Status: preliminary; translated from GR/EMBL/DBDB
A:Molecule type: DNA
A:Residues: 1-1777 <FAV>
A:Cross-references: EMBL:U041263; PIDN:AAC244.8.1; GSPDB:GN00020; CESP:T19D12.1
A:Experimental source: strain Bristol N2; clone T19D12
C:Genetics:
A:Gene: CESP:T19D12.1
A:Map position: 2
A:Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681/1

```

[illegible]

```

Db      399  PSTGCVPTSSKSSPPN-----SIPPTTIP-----GAPSTL 430
Oy      217  GONWKKIVLEQLDAKIEAEETRL-YVELKSSOPHLPELKKSVALLQPLNOSFI 275
Db      431  GSSSTIVSTTTIPS-----TPKASTLLSOSTPTSTPLVSS----- 469
Oy      276  QOCVOQTSSDMKVIATCTTTVTTSPPVTTVTSSSOSEKSIIVSGATAPRTVSOTLPLAG 335
Db      470  -----SSSSSSSTVTTSTITPTSTGCVPIST--SNQPMPS--TSNPPTPK--STVTAPSPIT 519
Oy      336  PVGAKAGVYVTHSGPPLPAALGGTTAGGLLOTSKPLVTSVANTVTYVLSLOPEKPVVSGTA 395
Db      520  GATSTASSTITTSAPTSPHSISPS--TMTSTVPVTSFSTASTTVPI---TVAAGOC 572
Oy      396  VTLSPAVTFGTSGAA 412
Db      573  KCQSNVAVAPELTSGTS 589

```

RESULT 11
F36791
C:Species: ictaluriid herpesvirus 1 (strain auburn 1)
A:Note: host Ictalurus punctatus (channel catfish)
C:Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 31-Jan-2000
A:Accession: F36791
R:Davidson, A.J.
submitted to Genbank, January 1992
A:Description: Channel catfish virus: a new type of herpesvirus.
A:Reference number: A36804
A:Accession: F36791
A:Molecule type: DNA
A:Residues: 1-670 <NAV>
A:Cross-references: GB:M75136; NID:g331209; PIDN:AAA88153.1; PID:g331260
R:Davidson, A.J.
Virology 186, 9-14, 1992
A:Title: Channel catfish virus: a new type of herpesvirus.
A:Reference number: A39447; M0ID:92087490; PMID:1727613
A:Contents: annotation
A:Note: neither protein nor nucleic acid sequence is given
C:Genetics:
A:Gene: 50
C:superfamily: period clock protein; EGF homology

Query Match	8.5%	Score 191	DB 2	Length 670
Best Local Similarity	24.8%	Pred. No. 0.00042		
Matches 110	Conservative 41	Mismatches 153	Indels 140	Gaps 22
OY	16	VSSGRLEPAPQVAAKAPNTTT-----IQFANLQLPGCVLIRKNSGGLMLVSPQOQVYT	70	
Db	248	VTTTPAMPAG--ANDTANITTAITPTGANDTANVTMPAGADTVTTTTPAM-----PT	297	
OY	71	RAETTSNITSRPAVPANQVICTVPPNSSOLIKKVAVTP-----VKRLAQIG	119	
Db	298	GANDTANITTT--ATPAGANDTANVTMPAGADTV--VTTTPMPAGANDTANVTMPAGST	353	
OY	120	TTVTVTTPVKPSSVOSVAVPT---SVVTTPGPKP--INTVTTLKPS--SLGASSTPNSBP	171	
Db	354	DTVVTTP-----AMPGATDVTVTTTPAMPGADIDTVTTTTPAMPGATDVTVTTT	405	
OY	172	NKRAENSAVOINISPTMLENVYKCKNEFLAMLIKLAGSGSPENGQVYKRLVEOLLDAK	231	
Db	406	PAKPGANGVTTTTPAM-----PAGAN-----	428	
OY	232	IEAEETFKLYELKSSPOPHLVPFLKKSVALROLLPNSOSFIQOCVOQT--SSDMVIAT	290	
Db	429	-----DTVTTTAPATPAGANDTANVTKPGATDVTVTT	461	
OY	291	CTT--TVTTSPTVTTTSSSSSEKSIIVSIGAT--APRTVSVQTLNPLAGPVGAKACAVTILH	347	
Db	462	ATVAKPTGATGVTTTTAKPGTANDTANVTKPGATGVTVTTT---AAKPGA--TCTVTIVA	516	
OY	348	SVGPAPATNGGTTA-----GGGLLOTSKPPLTVSVANTVTYVSLQDEKPYVSGTAATVL	398	

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 15.2925 Seconds

(Without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452
Perfect score: 2249
Sequence: 1 GLVTWKVAVSAPKRVSSGP.....VIGTPOIKLAPGVLSSQ 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR:*
2: PIR:*
3: PIR:*
4: PIR:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241.5	10.7	921	2	A48184
2	241.5	10.7	921	2	A45183
3	215	9.6	1275	2	T33369
4	209.5	9.3	528	2	I47141
5	204	9.1	2035	2	A40718
6	202	9.0	3020	2	A43932
7	200.5	8.9	1367	1	S48478
8	197.5	8.6	873	2	A47282
9	194	8.6	865	2	A47282
10	192	8.5	1777	2	T34369
11	191	8.5	670	2	F36791
12	190.5	8.5	825	2	T29634
13	186.5	8.3	873	2	F96615
14	185.5	8.2	3570	2	T45025
15	185	8.2	648	2	PC4395
16	183.5	8.2	2232	2	T34434
17	182.5	8.1	2187	2	T30826
18	181.5	8.1	662	2	A45185
19	179.5	8.0	867	2	T45463
20	179.5	8.0	1324	2	S52863
21	177.5	7.9	3507	2	T34513
22	175.5	7.8	796	2	T21460
23	174	7.7	549	2	C87719
24	173	7.7	1630	2	A53577
25	172.5	7.7	797	1	VGBEX1
26	172	7.6	886	2	S29605
27	171	7.6	1032	2	T45463
28	171	7.6	1032	2	T45463
29	170	7.6	1161	2	S57180

30	169.5	7.5	1151	2	T18535	high molecular mas
31	169	7.5	1367	2	S51959	hypothetical prote
32	167.5	7.4	725	2	A41258	a-agglutinin core
33	167	7.4	798	2	T34248	hypothetical prote
34	167	7.4	5376	2	T42215	zonadhesin - mouse
35	166	7.4	851	2	T22696	hypothetical prote
36	165	7.3	881	2	S56032	probable membrane
37	164.5	7.3	752	4	A57784	AMU1/MTG8 mutant f
38	164	7.3	1199	2	A40670	nuclear envelope p
39	164	7.3	1365	2	S14871	suppressor two of
40	163	7.2	896	2	T22061	hypothetical prote
41	162.5	7.2	1260	2	S60896	agglutinin-like pr
42	161.5	7.2	1832	2	T31113	mucin-like glycopr
43	161	7.2	2422	2	T12687	ALR protein homolo
44	160.5	7.1	780	2	A48143	HF-1 regulatory el
45	160	7.1	860	2	JC4566	chitinase (EC 3.2.

ALIGNMENTS

RESULT 1

A48184 transcription initiation factor IID 110k chain - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C:Accession: A48184

R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.

Proc. Natl. Acad. Sci. U.S.A. 90, 5896-5900, 1993

A>Title: The Drosophila 110-kDa transcription factor TFIID subunit directly interacts

A:Reference number: A48184; MUID:93317591; PMID:8327460

A:Accession: A48184

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-921 <KOK>

A:Cross-references: GB:563550; NID:9398432; PID:9398433

A:Experimental source: embryo nuclear extract

A>Note: Sequence extracted from NCBI Backbone (NCBI:134863, NCBI:134864)

C:Genetics:

A:Gene: FlyBase:Tafl10

A:Cross-references: FlyBase:FBgn0010280

C:Keywords: transcription initiation

Query Match	10.7%	Score 241.5;	DB 2;	Length 921;
Best Local Similarity	24.8%	Pred. No. 5.6e-07;		
Matches 112;	Conservative 65;	Mismatches 163;	Indels 111;	Gaps 17;
OY	25	PQIVAVKAPNTTTOFPANLQLPPTVLIKNSGPIML	-----VSPQQTVAEFTTS	76
DB	133	POSPITLSTINTGOTPA-----LLVKTDNGQLRVGTTGPPVTQTITTSNNS	184	
OY	77	NITSPAVPANTTTOFPANLQLPPTVLIKNSGPIML	-----LIKKVAVTPVKKLAQIGTVVTVTP	127
DB	185	NTSTNTNPTTTO--IRLOTVAASMTTNTSNIVASVSGVANSOPHLLQLNAQ	243	
OY	128	KPSVQSVAVPTSVVTVPGKPLNTVTTLKPSSSLGASSTPSNEPLAENSAVAQINLSP	187	
DB	244	APQIPQITQITQITPAQSOQOOVNVVSSAGTAAVASTTA-----ATT	287	
OY	188	TMLENVK-KCKNPLAMLIKACSGSQSPDMGONKVLVEQLLDKLEAEFTRLYELK	246	
DB	288	TQGGTEKCKKFLANLEL--STREPKVKKVNTLIQELVANVEEFCDLERILN	345	
OY	247	SSPQHLVPLFKKSVVALROL-----LPSQSFIO---	276	
DB	346	ASPPQCLIGFLAKSLPRLQALYKELVIEGIRKPPQHVGLAGLSQQLPFIQAOIRIG	405	
OY	277	---QCVQOISSDVAIATCTTPTTSPVVTTVSSSQSEKSIIVSGATAPRTVS---VQT	329	
DB	406	PSQTTTIGOTGVRI--TPNALGTPRPTIGHTITSKQPN--IRLPAPRLVNTGIRI	460	
OY	330	LNPAGPVGAGAGVVTLSHVGPTAATGTTAGTGLQTSKRLVTVSVANTVTVVSLQPEKP	389	

This Page Blank (uspto)

```

Db 414 PTATSGIRATLPTVLAPRLPOP-----PONPTNIQ---NFOLPGMVLVRSNGQL 464
QY 62 LVSPQOQYTR-----AETTSNITSRAVPANPQVTKICTVPNSSQLIKKVAATPVKKL 115
Db 465 MI-PQOALAQOQAHAPQOTMAPRPAFTPSAPPVQISTVQAPGPTIAR-QVTP----- 518
QY 116 AQIGTVTVTVKPPSSVQSVAVPTSVTVTPGKPLNTVT--TLKPSISGASS-----TPS 168
Db 519 -----TTIKQY---SQOQTVQPSKTLQSPGVQPOLVGGAAQTASLGTAIVQGTIPQ 571
QY 169 NE-PNLKAENSAVQINLSPTMLENVKKCKNFLAMLIKIACSGSQSPMGQNVKKLVEQL 227
Db 572 RTVPQATTTSSAATE-----TMENVKKCKNFLSTLIKIASGKQSTETJANVKELVQNL 625
QY 228 LDKIEAEFEETRKLYVELKSSPQPHLVPLFKKSVVALRQLLPNSQSTIQCVQO----TS 283
Db 626 LDGRIEADFTSRLYRELNSSPOPVLVPLFKRSLPALRQLTPDSAAFTQSSQOQPPPTS 685
QY 284 SDWVIATCTTVTVTPSVVTVTVSSSQSEKSTIVSGATAPRTVSVQTLNPLAGPVGAKAGV 343
Db 686 Q-----ATTALNAVVLSSSVQRTAGTKAATVTSALQPVVLSL----- 722
QY 344 VTLHSVGPAAATGTTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTLSPAV 403
Db 723 -----TQPVGVGVGKQGOPTPLVIO-----QPKP----- 747
QY 404 TFGETSGAICLPsvkPvsvfcmDhICKPVI--GTPV-----QIKLAOPG 446
Db 748 -----GALIOINLPQPV-----PVVKPVALPPTKALSAVSAQAAAAQKNKKEPG 792

```

Search completed: February 16, 2003, 22:04:22
 Job time : 20.3897 secs


```
Db 530 LCKIEAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSAAFIOOQOQPPPTS 589
OY 284 SDWVATCTTTVTTSPPVTTTSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKGV 343
Db 590 Q-----ATTALTAVALSSVQRTAGKTAATVTSALQPPVLSL----- 626
OY 344 VTLHSVGPAAAGCTTAGGGLLOTSKPLVTSVANTVTVTSLOPEKRVSGTAVTSLPVA 403
Db 627 -----TOPTOVGKOGQPPPLVIO-----QPPKP----- 651
OY 404 TFGTSGAALCLPSVAVSFCMDHICKPVI--GTPV-----OIKLAORG 446
Db 652 -----GALIQLNPLQPV-----PVVKAVALPGLTKALSAVSAQAAAQKNKLEBG 696

RESULT 13
US-09-724-676-63923
; Sequence 63923, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 2000-11-28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63923
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63923

Query Match
Best Local Similarity 32.6%; Score 507.5; DB 5; Length 775;
Best Local Similarity 32.9%; Pred. No. 1.3e-20;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

OY 13 PKVSSG-----PRLAPQIVAVKAPRTTITQPPANIQLPBGVILKNSGPI 61
Db 382 PTAISGIRATLPTVLARLP-----PONPTNIQ---NFOLPPGVAVLSSENGOLL 432
OY 62 LVSPQOVTNR-----AETTSNITSRAVPANPQVTKICTVPSNSSLKIKKVAVTPVK 115
Db 433 MI-PQOALQMOQAHAQOQTMAAPPATPTSAAPPVQISTVOAGPTPIAR-QVTP----- 486
OY 116 AQIGTVVTVTPKPSVQSAVPTSVVTVTPGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 487 -----TTIKQV---SQAQTTVQPSATLQSPGVQPLVGGAAQRTASLGATAVQGTG 539
OY 169 NE-PNLKAENSAVQINLSPTMLENKKCKNFLAMLIKILACSGSGSPENQONVKLIVEOL 227
Db 540 RTVPGATTTSSATE-----TWNENKCKKNFLSTLIKILASSGKOSTETAAVKEIYQNL 593
OY 228 LQAKIAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSAAFIOOQOQPPPTS 283
Db 594 LQAKIAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSAAFIOOQOQPPPTS 653
OY 284 SDWVATCTTTVTTSPPVTTTSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKGV 343
Db 654 Q-----ATTALTAVALSSVQRTAGKTAATVTSALQPPVLSL----- 690
OY 344 VTLHSVGPAAAGCTTAGGGLLOTSKPLVTSVANTVTVTSLOPEKRVSGTAVTSLPVA 403
Db 691 -----TOPTOVGKOGQPPPLVIO-----QPPKP----- 715
OY 404 TFGTSGAALCLPSVAVSFCMDHICKPVI--GTPV-----OIKLAORG 446
Db 716 -----GALIQLNPLQPV-----PVVKAVALPGLTKALSAVSAQAAAQKNKLEBG 760

RESULT 14
US-09-724-676A-63923
; Sequence 63923, Application US/09724676A
; GENERAL INFORMATION:
```

```
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 2000-11-28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63923
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63923

Query Match
Best Local Similarity 22.6%; Score 507.5; DB 5; Length 775;
Best Local Similarity 32.9%; Pred. No. 1.3e-20;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

OY 13 PKVSSG-----PRLAPQIVAVKAPRTTITQPPANIQLPBGVILKNSGPI 61
Db 382 PTAISGIRATLPTVLARLP-----PONPTNIQ---NFOLPPGVAVLSSENGOLL 432
OY 62 LVSPQOVTNR-----AETTSNITSRAVPANPQVTKICTVPSNSSLKIKKVAVTPVK 115
Db 433 MI-PQOALQMOQAHAQOQTMAAPPATPTSAAPPVQISTVOAGPTPIAR-QVTP----- 486
OY 116 AQIGTVVTVTPKPSVQSAVPTSVVTVTPGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 487 -----TTIKQV---SQAQTTVQPSATLQSPGVQPLVGGAAQRTASLGATAVQGTG 539
OY 169 NE-PNLKAENSAVQINLSPTMLENKKCKNFLAMLIKILACSGSGSPENQONVKLIVEOL 227
Db 540 RTVPGATTTSSATE-----TWNENKCKKNFLSTLIKILASSGKOSTETAAVKEIYQNL 593
OY 228 LQAKIAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSAAFIOOQOQPPPTS 283
Db 594 LQAKIAEFTSKLYRELNSPOPIVLPFLKSLPALROLTPDSAAFIOOQOQPPPTS 653
OY 284 SDWVATCTTTVTTSPPVTTTSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKGV 343
Db 654 Q-----ATTALTAVALSSVQRTAGKTAATVTSALQPPVLSL----- 690
OY 344 VTLHSVGPAAAGCTTAGGGLLOTSKPLVTSVANTVTVTSLOPEKRVSGTAVTSLPVA 403
Db 691 -----TOPTOVGKOGQPPPLVIO-----QPPKP----- 715
OY 404 TFGTSGAALCLPSVAVSFCMDHICKPVI--GTPV-----OIKLAORG 446
Db 716 -----GALIQLNPLQPV-----PVVKAVALPGLTKALSAVSAQAAAQKNKLEBG 760

RESULT 15
US-09-724-676-63926
; Sequence 63926, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 2000-11-28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63926
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63926

Query Match
Best Local Similarity 22.6%; Score 507.5; DB 5; Length 807;
Best Local Similarity 32.9%; Pred. No. 1.3e-20;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

OY 13 PKVSSG-----PRLAPQIVAVKAPRTTITQPPANIQLPBGVILKNSGPI 61
```

```

OY      13  PKVSSG-----PRLPAPQIVANKANNTTIOPRALQIPRGIVLILKNSGRLM  61
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      318  PATTSSGIRATILTPVLPVLRPOP-----PONPNIIQ--NFOLPBGAVLVRSENGOLL  368

OY      62  LVSPQOQVTR-----AETTSNITSRPAVPAPQTVICTVPSNSQLIKKVAIVPVKVL  115
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      369  MI-PQOALAQMOQAHAQOPQTMARPRATPRISAPRYQISTVQAPGPTIAR-QVTR-----  422

OY      116  AOGITVVTVTPKPSVSVAVPISVTVTPGKPLNTYT--TLKPSISGASS-----TPS  168
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      423  ----TIIIKOV---SQAQTVQPSATILQSPFCVQPOLVIGGAQAQIASLGTATAVOTGTFQ  475

OY      169  NE-PYLKAKENSAAVQINLSPTMLEVNVKKCKNFLMLIKIACGSGSPSPKGVAVKLYEVL  227
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      476  RTVPATATTSKATE-----TMEVVKCKCKNFLSTLIKIASSGKSTETAAVVKELVOML  529

OY      228  LDKATIEAEFTKRLVELKSSPQPLVLPFLKSSVALROLRLPNSOSFTIOQCVQ-----TS  283
           |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      530  LDGKIEADEFTSRILRELNSSPQPLVLPFLKSLRLALRQLTDSNAFTIOSSQOQPPPPIS  589

OY      284  SDMVATATCTTVTISPVVTTTVSSSQSEKSIIVSGATAPRVSYOTLNLPLAGPVGAKAGV  343
           ||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      590  Q-----ATTALITAVVLISSSVQRTAGKTAATVTSALAPVLSL-----  626

OY      344  VTLHSVGPFAANGCTAGGLQTSKPLVTSVAANVTTVTSLOPEKPVVSGTAVTISLRAV  403
           |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      627  ----TQPIQVGVGKQOQTPPLVIO-----QPPKP-----  651

OY      404  TFGTSGAICLPSKPVVSEFCMDICKRVI--GTPV-----QIKLAPG  446
           |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      652  ----GALIQINPLIQV-----PVVKPRAVLPGRIKALISAVSAQAQAQKNNLKEBG  696

RESULT 12
US-09-724-676A-63920
: Sequence 63920, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OR INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724, 676A
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63920
: LENGTH: 711
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-63920

Query Match          22.6%: Score 507.5; DB 5; Length 711;
Best Local Similarity 32.9%: Pred. No. 1,le 20;
Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

OY      13  PKVSSG-----PRLPAPQIVANKANNTTIOPRALQIPRGIVLILKNSGRLM  61
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      318  PATTSSGIRATILTPVLPVLRPOP-----PONPNIIQ--NFOLPBGAVLVRSENGOLL  368

OY      62  LVSPQOQVTR-----AETTSNITSRPAVPAPQTVICTVPSNSQLIKKVAIVPVKVL  115
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      369  MI-PQOALAQMOQAHAQOPQTMARPRATPRISAPRYQISTVQAPGPTIAR-QVTR-----  422

OY      116  AOGITVVTVTPKPSVSVAVPISVTVTPGKPLNTYT--TLKPSISGASS-----TPS  168
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      423  ----TIIIKOV---SQAQTVQPSATILQSPFCVQPOLVIGGAQAQIASLGTATAVOTGTFQ  475

OY      169  NE-PYLKAKENSAAVQINLSPTMLEVNVKKCKNFLMLIKIACGSGSPSPKGVAVKLYEVL  227
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      476  RTVPATATTSKATE-----TMEVVKCKCKNFLSTLIKIASSGKSTETAAVVKELVOML  529

OY      228  LDKATIEAEFTKRLVELKSSPQPLVLPFLKSSVALROLRLPNSOSFTIOQCVQ-----TS  283
           |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      530  LDGKIEADEFTSRILRELNSSPQPLVLPFLKSLRLALRQLTDSNAFTIOSSQOQPPPPIS  589

OY      284  SDMVATATCTTVTISPVVTTTVSSSQSEKSIIVSGATAPRVSYOTLNLPLAGPVGAKAGV  343
           ||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      590  Q-----ATTALITAVVLISSSVQRTAGKTAATVTSALAPVLSL-----  626

OY      344  VTLHSVGPFAANGCTAGGLQTSKPLVTSVAANVTTVTSLOPEKPVVSGTAVTISLRAV  403
           |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      627  ----TQPIQVGVGKQOQTPPLVIO-----QPPKP-----  651

OY      404  TFGTSGAICLPSKPVVSEFCMDICKRVI--GTPV-----QIKLAPG  446
           |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      652  ----GALIQINPLIQV-----PVVKPRAVLPGRIKALISAVSAQAQAQKNNLKEBG  696

```

Query Match 23.2%; Score 521; DB 5; Length 837;
 Best Local Similarity 34.0%; Pred. No. 2,4e-21;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

13 PRVSSG-----RLRPAQIVAAKAPRTTIOFPANIQLPBGIVLKSNGPLM 61
 DB 414 PRATTSIGIRATLPTVLAPELPDP-----PONPINIQ--NFQLPBGIVLSENGOLL 464
 QY 62 LVSPOQVTR-----AETTSNITSRAVPANQVKTICVTPNSSSOLIKKAVTPVKKL 115
 DB 465 MI-PQOALQMOQAHAQOQTMAAPRATPTSPAPVOISTVQAPGPITAR-QVTP----- 518
 QY 116 AQIGTTVTTVPKPSVQSAVPTSVVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
 DB 519 -----TTIIKQV--SOAQTVQPSATLQRSPOGVQPOLVGAQAQTASLGATAVQGTFO 571
 QY 169 NE-PNKAENSAVQINLSPTMLENKKCKNPLAMLIKALSSGSSQSPKQONKAYLEOL 227
 DB 572 RIVPGATTTSSATE-----TMEVKKCKNPLSTLIKILASSGOSTETAAVKELOVNL 625
 QY 228 LDKIAEETTRKLYELKSSPOPHLVPLFKRSVALROLPLPSOSFIOOCVQO-----TS 283
 DB 626 LDKIAEDFTSLYRELNLSPOPIVPLFKRSLPALROLTPDSAAFIOOSQOQPPPPPS 685
 QY 284 SDNVIACTTTVTTSPVTTVSSSSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
 DB 686 Q-----ATTALTAVLVSSVQRTAGTAATVTSALQPPVLST----- 722
 QY 344 VTLHSGPAAAGCTAGTACGLQTSKPLVTSVANTVTVTSLOPEKPVSGTAVTSLPAPV 403
 DB 723 -----TQPTQVGVKQOGQPTPLVQ-----QPPKP--GALIRPQV 756
 QY 404 TFEETSGAAILCPSPVSVFCMDHICKPVIQPIKLT--AQPGVYL 449
 DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQOIQLNPLOPVV 792

RESULT 8
 US-09-724-676A-63924
 ; Sequence 63924, Application US/09724676A
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724, 676A
 ; CURRENT FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 63924
 ; LENGTH: 837
 ; TYPE: PRV
 ; ORGANISM: Homo sapiens
 US-09-724-676A-63924

Query Match 23.2%; Score 521; DB 5; Length 837;
 Best Local Similarity 34.0%; Pred. No. 2,4e-21;
 Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

13 PRVSSG-----RLRPAQIVAAKAPRTTIOFPANIQLPBGIVLKSNGPLM 61
 DB 414 PRATTSIGIRATLPTVLAPELPDP-----PONPINIQ--NFQLPBGIVLSENGOLL 464
 QY 62 LVSPOQVTR-----AETTSNITSRAVPANQVKTICVTPNSSSOLIKKAVTPVKKL 115
 DB 465 MI-PQOALQMOQAHAQOQTMAAPRATPTSPAPVOISTVQAPGPITAR-QVTP----- 518
 QY 116 AQIGTTVTTVPKPSVQSAVPTSVVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
 DB 519 -----TTIIKQV--SOAQTVQPSATLQRSPOGVQPOLVGAQAQTASLGATAVQGTFO 571
 QY 169 NE-PNKAENSAVQINLSPTMLENKKCKNPLAMLIKALSSGSSQSPKQONKAYLEOL 227
 DB 572 RIVPGATTTSSATE-----TMEVKKCKNPLSTLIKILASSGOSTETAAVKELOVNL 625

QY 228 LDKIAEETTRKLYELKSSPOPHLVPLFKRSVALROLPLPSOSFIOOCVQO-----TS 283
 DB 626 LDKIAEDFTSLYRELNLSPOPIVPLFKRSLPALROLTPDSAAFIOOSQOQPPPPPS 685
 QY 284 SDNVIACTTTVTTSPVTTVSSSSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
 DB 686 Q-----ATTALTAVLVSSVQRTAGTAATVTSALQPPVLST----- 722
 QY 344 VTLHSGPAAAGCTAGTACGLQTSKPLVTSVANTVTVTSLOPEKPVSGTAVTSLPAPV 403
 DB 723 -----TQPTQVGVKQOGQPTPLVQ-----QPPKP--GALIRPQV 756
 QY 404 TFEETSGAAILCPSPVSVFCMDHICKPVIQPIKLT--AQPGVYL 449
 DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQOIQLNPLOPVV 792

RESULT 9
 US-09-724-676-63929
 ; Sequence 63929, Application US/09724676
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724, 676
 ; CURRENT FILING DATE: 2000-11-28
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 63929
 ; LENGTH: 679
 ; TYPE: PRV
 ; ORGANISM: Homo sapiens
 US-09-724-676-63929

Query Match 22.6%; Score 507.5; DB 5; Length 679;
 Best Local Similarity 32.9%; Pred. No. 1,1e-20;
 Matches 157; Conservative 58; Mismatches 121; Indels 141; Gaps 20;

13 PRVSSG-----RLRPAQIVAAKAPRTTIOFPANIQLPBGIVLKSNGPLM 61
 DB 286 PRATTSIGIRATLPTVLAPELPDP-----PONPINIQ--NFQLPBGIVLSENGOLL 336
 QY 62 LVSPOQVTR-----AETTSNITSRAVPANQVKTICVTPNSSSOLIKKAVTPVKKL 115
 DB 337 MI-PQOALQMOQAHAQOQTMAAPRATPTSPAPVOISTVQAPGPITAR-QVTP----- 390
 QY 116 AQIGTTVTTVPKPSVQSAVPTSVVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
 DB 391 -----TTIIKQV--SOAQTVQPSATLQRSPOGVQPOLVGAQAQTASLGATAVQGTFO 443
 QY 169 NE-PNKAENSAVQINLSPTMLENKKCKNPLAMLIKALSSGSSQSPKQONKAYLEOL 227
 DB 444 RIVPGATTTSSATE-----TMEVKKCKNPLSTLIKILASSGOSTETAAVKELOVNL 497
 QY 228 LDKIAEETTRKLYELKSSPOPHLVPLFKRSVALROLPLPSOSFIOOCVQO-----TS 283
 DB 498 LDKIAEDFTSLYRELNLSPOPIVPLFKRSLPALROLTPDSAAFIOOSQOQPPPPPS 557
 QY 284 SDNVIACTTTVTTSPVTTVSSSSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
 DB 558 Q-----ATTALTAVLVSSVQRTAGTAATVTSALQPPVLST----- 594
 QY 344 VTLHSGPAAAGCTAGTACGLQTSKPLVTSVANTVTVTSLOPEKPVSGTAVTSLPAPV 403
 DB 595 -----TQPTQVGVKQOGQPTPLVQ-----QPPKP--GALIRPQV 619
 QY 404 TFEETSGAAILCPSPVSVFCMDHICKPVI--GPV-----QIKLAQPG 446
 DB 620 -----GALIQNLPLQPV-----PVKPAVILPGRKALSASVSAQAAAQKNKLEKG 664

RESULT 10

QY 404 TFGETSGAICLPKRVNSECMDHICKPVIGTPVQIKL--AQGPVL 4493
| : | : | : | : | : | :
Db 629 TLQT-----PMVALRQH-NRIMLTPOQIQLNPLQRPVV 6644

RESULT 2
US-09-724-676A-63927
; Sequence 63927, Application US/09724676A

```

: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63927
: LENGTH: 709
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-724-676A-63927

```

Query Match	23.2%	Score 521;	DB 5;	Length 709;
Best Local Similarity	34.0%	Pred. No. 2	1e-21;	
Matches 159;	Conservative	61;	Mismatches 128;	Indels 120;
				Gaps 20

```

QY 13 PRKSSG-----PRLEPAQIYAIVAKAPNTTITQIPANLOIPGVLTIAINSQGLM 61
Db 286 PLATISGRATLTVTYLAIRLPQ-----PQNTNIQ-----NQLPFGMTLVISENGQL 3366
QY 62 LVSPQQTIVTR-----AETTSNITSRPDVPANPDQVIGICVPMSSSOLIKRAVTPYKTL 115
Db 337 MI-FOQALAQOAOAHQAPOTTPMAREPTEPSAPPVISTVQAFGEPIIAR-QVTP----- 3906
QY 116 AQITVTYVTPKRSVOSAVAPRVSUTVTPGKRLNVT--TLKPSLSGASS-----TPS 168
Db 391 -----TTIIKQV-----SQAQITVQSPATIQRSNGVOPOLVIGAAQAFASLGATATVQCTPO 4433
QY 169 NE-ENLEAENSAVQOINLSPTMLEENKCKNFIAMLITKLACSGSOSPEMQONKYLEOL 2277
Db 444 RTVGATTTSSAATE-----TMENVKCKKNFISTILIKLASSGQSTETANAKELQNL 497
QY 228 LDATIEAEFPRKRIYVELKSSPOHLVPLFKSVYALRQLLPMSQSTIQOCVQO-----TS 283
Db 498 LDGIEADEDFSTRILYRELNSPQVLPVPLFKSRIPALRQLTPPSAAETIQOSQOOPPEPTS 5577
QY 284 SDMWIATCTTWTTPSPVVTYVSSSOSSEKSIIVSGATAPRTVSQVLEMLPAGPVAKAGV 3434
Db 558 Q-----ATPALAVAVLSSVQGTACKTAAATVISAQPPVLSL-----594
QY 344 VTLHSVGPTAATGGTGTAGTGLQTSKPLVTVSVANTVTVTSILOPEKPPVSGTAVTLSPAV 4033
Db 595 -----TOPTQVGVGQOGQPTPLVQO-----QBPKP-----GALIRPPQV 6288
QY 404 TFGETSGAICLPKRVKPVVSCMHICKPVLIGVNOIKL--AQRPVL 449
Db 629 TLITGT-----PVALARQRP-NRIKITTPQOIQLOMLPDRVPV 664

```

```

RESULT 3
US-09-724-676-63930
: Sequence 63930, Application US/09724676
:
GENERAL INFORMATION:
:
APPLICANT: CompuGen LTD
:
TITLE OF INVENTION: Variants of alternative splicing
:
FILE REFERENCE: 129181.4 CompuGen
:
CURRENT APPLICATION NUMBER: US/09/724.676
:
CURRENT FILING DATE: 2000-11-28
:
NUMBER OF SEQ ID NOS: 97222
:
SOFTWARE: PatentIn version 3.2
:
SEQ ID NO 63930
:
LENGTH: 741
:
TYPE: PRT

```

```

: ORGANISM: Homo sapiens
US-09-724-676-63930

Query Match      23.2%, Score 521; DB 5; Length 741;
Best Local Similarity 34.0%, Pred. No. 2,1e-21;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

OY 13 PPKVSSG-----PRLPAPQIVAVKAPANTTIOEPANLOLPKPGVLIKSSGPLM 61
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 318 PTTATSSGRATLPTVLAPRLPDP-----PONPNINIO---NFQLRPGKAVLYRSENGQL 368
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 62 LVSPQGYTR-----AETTSNITSRPAPVPANPOTVKTICTVPNSSSOLIKKVAATPVKKL 115
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 369 MI-PQOALAQMOAQOAHQAPOTTMARPRATPTSAPRVOISTVOAGPCTPIIAR-QVTP---- 422
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 116 AQTGTYTTPPKPSSVQSANPTSVVTVTPCKPLNTYT--TLKSSILGASS-----TPS 168
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 423 ----TTIIKQV---SQADTVQPSATLDRSPGVQPOLVILGAAQVASIGTAAVQGTPTQ 475
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 169 NE-PMLKENSAAVQOINISPTMLEVKKCKNMLMIKLKAGSSQSPENQGVNKKLVEOL 227
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 476 RVEPATITSSAATP-----TMEVKKCKNLTSLIKLASSGKSTETAAVAVKELVQNL 529
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 228 LDKATLAEAEFFRKRLKVELKSSPQPHLPVPLKKSVALKQLLPNSQSFIOQCYQO----TS 283
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 530 LDGKLEADPFTSRRLRELNSPQPLVPEFLKRSPLALROLTPDSAAFTIQSSQOQPPPTPS 589
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 284 SDNVATCTTPTVTSPPVTTTVSSSQSEKSIIVSGATPRTVSYOTLNDPLAGPVAKAGV 343
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 590 Q-----ATTALRAVVLSSSVQKAGTAAATVTSALQPPVLSL----- 626
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 344 VTLHSGEPTATGGTTAGTGLQTSKRLPVTSVANTVTVTSLOPEKPVVSGAVTSLPAPV 403
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 627 -----TOPVQGVGKQGPPTPLVIO-----QPRKR-----GALRRPQV 660
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 404 TEGESGAATCLPSKVPYVSEFMDHCKRPVGTQIKL--AQGPVVL 449
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 661 TLTQI-----PMVALRKQPI-NRIMLTTPQOILNPLQPVV 696
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 4
US-09-724-676A-63930
: Sequence 63930, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: CompuGen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 CompuGen
: CURRENT APPLICATION NUMBER: US/09/724.676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63930
: LENGTH: 741
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-63930

Query Match      23.2%, Score 521; DB 5; Length 741;
Best Local Similarity 34.0%, Pred. No. 2,1e-21;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

OY 13 PPKVSSG-----PRLPAPQIVAVKAPANTTIOEPANLOLPKPGVLIKSSGPLM 61
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 318 PTTATSSGRATLPTVLAPRLPDP-----PONPNINIO---NFQLRPGKAVLYRSENGQL 368
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 62 LVSPQGYTR-----AETTSNITSRPAPVPANPOTVKTICTVPNSSSOLIKKVAATPVKKL 115
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 369 MI-PQOALAQMOAQOAHQAPOTTMARPRATPTSAPRVOISTVOAGPCTPIIAR-QVTP---- 422
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 116 AQTGTYTTPPKPSSVQSANPTSVVTVTPCKPLNTYT--TLKSSILGASS-----TPS 168
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 423 ----TTIIKQV---SQADTVQPSATLDRSPGVQPOLVILGAAQVASIGTAAVQGTPTQ 475
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 169 NE-PMLKENSAAVQOINISPTMLEVKKCKNMLMIKLKAGSSQSPENQGVNKKLVEOL 227
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 476 RVEPATITSSAATP-----TMEVKKCKNLTSLIKLASSGKSTETAAVAVKELVQNL 529
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 228 LDKATLAEAEFFRKRLKVELKSSPQPHLPVPLKKSVALKQLLPNSQSFIOQCYQO----TS 283
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 530 LDGKLEADPFTSRRLRELNSPQPLVPEFLKRSPLALROLTPDSAAFTIQSSQOQPPPTPS 589
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 284 SDNVATCTTPTVTSPPVTTTVSSSQSEKSIIVSGATPRTVSYOTLNDPLAGPVAKAGV 343
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 590 Q-----ATTALRAVVLSSSVQKAGTAAATVTSALQPPVLSL----- 626
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 344 VTLHSGEPTATGGTTAGTGLQTSKRLPVTSVANTVTVTSLOPEKPVVSGAVTSLPAPV 403
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 627 -----TOPVQGVGKQGPPTPLVIO-----QPRKR-----GALRRPQV 660
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 404 TEGESGAATCLPSKVPYVSEFMDHCKRPVGTQIKL--AQGPVVL 449
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 661 TLTQI-----PMVALRKQPI-NRIMLTTPQOILNPLQPVV 696
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 CompuGen
: CURRENT APPLICATION NUMBER: US/09/724,676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63930
: LENGTH: 741
: TYPE: PR1
: ORGANISM: Homo sapiens
US-09-724-676A-63930

Query Match          23.2%, Score 521; DB 5; Length 741;
Best Local Similarity 34.0%, Pred. No. 2,1e-21;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY      13  PPKVSSG-----PRLPAPIYAVKAPANTTTTTPQPNALQLPQGVTLIKNSGPILM 61
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      318  PTTATTSIGIRATLPTPTVLAPRLPQP-----PONPTNID--NFQILPQGVTLVRSNGQLL 368

QY      62  LVSPQOQVTR-----AETTSNITSRPAPVAPNAPQVTKICTYVNSSSOLIKKVAATVPVKKL 115
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      369  MI-PQQAALAQQAQNAHPQRTTMARPRATPPISAPRVQISTYQAPCPITLIAR-QVTRP----- 422

QY      116  AQIGTTVVTYPRKSSVSVSAVPTSVTVTPGKPLNTVT--TLKPSIGASS-----TPS 168
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      423  ----TTTIKQV----SQAOCTVQPSATLQRSQVQPOLVILGAGCAQATASIGTATATVQGTTPQ 475

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:00:19 : Search time 18.3897 Seconds
(Without alignments)
2076.603 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452
Perfect score: 2249
Sequence: 1 GLVTKVAPVAPKXSSGP.....VIGTPVQIKLQGPVLISQP 452

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 429898 seqs, 84487048 residues

Total number of hits satisfying chosen parameters: 429898

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCF_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	521	23.2	709	US-09-724-676-63927	Sequence 63927, A
2	521	23.2	709	US-09-724-676A-63927	Sequence 63927, A
3	521	23.2	741	US-09-724-676A-63930	Sequence 63930, A
4	521	23.2	741	US-09-724-676A-63930	Sequence 63930, A
5	521	23.2	805	US-09-724-676-63921	Sequence 63921, A
6	521	23.2	805	US-09-724-676A-63921	Sequence 63921, A
7	521	23.2	837	US-09-724-676-63924	Sequence 63924, A
8	521	23.2	837	US-09-724-676A-63924	Sequence 63924, A
9	507.5	22.6	679	US-09-724-676-63929	Sequence 63929, A
10	507.5	22.6	679	US-09-724-676A-63929	Sequence 63929, A
11	507.5	22.6	711	US-09-724-676-63920	Sequence 63920, A
12	507.5	22.6	711	US-09-724-676A-63920	Sequence 63920, A
13	507.5	22.6	775	US-09-724-676-63923	Sequence 63923, A
14	507.5	22.6	775	US-09-724-676A-63923	Sequence 63923, A
15	507.5	22.6	807	US-09-724-676-63926	Sequence 63926, A
16	507.5	22.6	807	US-09-724-676A-63926	Sequence 63926, A
17	507.5	22.6	925	US-09-724-676-63928	Sequence 63928, A
18	507.5	22.6	925	US-09-724-676A-63928	Sequence 63928, A
19	507.5	22.6	957	US-09-724-676-63931	Sequence 63931, A
20	507.5	22.6	957	US-09-724-676A-63931	Sequence 63931, A
21	507.5	22.6	1021	US-09-724-676-63922	Sequence 63922, A
22	507.5	22.6	1021	US-09-724-676A-63922	Sequence 63922, A
23	507.5	22.6	1053	US-09-724-676-63925	Sequence 63925, A
24	507.5	22.6	1053	US-09-724-676A-63925	Sequence 63925, A
25	195.5	8.7	5179	US-60-438-735-151	Sequence 151, App
26	191	8.5	2108	PCT-US02-33723-2	Sequence 2, App11

27	191	8.5	2135	6	US-10-288-798-9	Sequence 9, App1
28	191	8.5	2382	6	US-10-196-935A-2	Sequence 2, App1
29	191	8.5	2382	6	US-10-293-017-48	Sequence 48, App1
30	191	8.5	2382	6	US-10-052-648A-40	Sequence 40, App1
31	191	8.5	2382	6	US-10-293-071-48	Sequence 48, App1
32	186.5	8.3	5374	6	US-10-028-248A-75	Sequence 75, App1
33	186.5	8.3	5374	6	US-10-107-782-75	Sequence 75, App1
34	181.5	8.1	1270	1	PCT-US02-38445-39	Sequence 39, App1
35	180.5	8.0	1322	6	US-10-218-140-4828	Sequence 4828, App
36	173.5	7.7	2781	6	US-10-263-929-122	Sequence 122, App
37	173.5	7.7	2870	5	US-09-655-160-15	Sequence 15, App1
38	173.5	7.7	3178	5	US-09-655-160-4	Sequence 4, App1
39	171.5	7.6	2117	6	US-10-120-801-63	Sequence 63, App1
40	171.5	7.6	2126	6	US-10-052-648A-39	Sequence 39, App1
41	168.5	7.5	5935	6	US-10-243-243A-8	Sequence 8, App1
42	168	7.5	2601	6	US-10-028-248A-76	Sequence 76, App1
43	168	7.5	2601	6	US-10-107-782-76	Sequence 76, App1
44	167.5	7.4	2971	6	US-10-218-140-1990	Sequence 190, App
45	167	7.4	5376	6	US-10-028-248A-74	Sequence 74, App1

ALIGNMENTS

RESULT 1
US-09-724-676-63927
Sequence 63927, Application US/09724676
GENERAL INFORMATION:
APPLICANT: CompuGen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181, 4 CompuGen
CURRENT APPLICATION NUMBER: US/09/724, 676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: Patentin version 3.2
SEQ ID NO 63927
LENGTH: 709
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-63927

Query Match 23.2% Score 521: DB 5: Length 709:

Best Local Similarity 34.0% Pred. No. 2.1e-21:
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY	13	PKVSSG-----PRLPADQIVAKPNTTTFQPNLQIPPTVLKNSGPI 61
DB	286	PTATSGIRATLPTVLARLP-----PQNTPTNIO---NFQLPFGMTLVRENGQL 336
QY	62	LVSPOQIVTR-----AETTSNITSRAVPANPQTKICTVPSNSQILKRVAVTPVK 115
DB	337	MI-POQALQOQAOAHAPQPTMAPRPATPTSPAPVQISTVQAPCTPIAR-QVTP--- 390
QY	116	AGIGTTVTTPKPSVQSAVNPVSVTTPGKPLNTV---TLKPSISGASS-----TPS 168
DB	391	-----TTIKOV-----SQQTIVQPSKTLQDSPEVQDVLGAAQTAASGTAVTGTPO 443
QY	169	NE-PULKAENSAVQINSPMLLENVKKCKNFMILKACSGSPQMGONVKKLEQL 227
DB	444	RTVEPATITSSAHE-----TMEVKKCKNFTSLTILKASSGOSTETANVEIYNL 497
QY	228	LDATKEAEFTFRKLYELKSSPQHLVPLFKSVYALQQLLPNSQSFQOCVQO---TS 283
DB	498	LDGKLEADFTSRILYRELNSPQPLVPLFKSLPALQLTPDSAAFTQSQDQPPPTPS 557
QY	284	SDMYATATCTVTTSPPVTTVSSSQSEKSTIVGATAPRTVSVQTLNPLGPGAKAGV 343
DB	538	O-----ATFALTRAVLVSSSVQRTAGTATVTSALQPVLSL-----594
QY	344	VTLSVGPATATGCTAGTGLQTSKPLVTSVANTVTVTSIQPEKPVYSGTAVTLSPAV 403
DB	595	-----TQPTQVGVGKGQPTPLVQO-----QPKRP-----GALIRPQV 628

This Page Blank (uspto)

;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
;; TITLE OF INVENTION: THEROPO
;; FILE REFERENCE: CLO00173
;; CURRENT APPLICATION NUMBER: US/60/173,464
;; CURRENT FILING DATE: 1999-12-29
;; NUMBER OF SEQ ID NOS: 30269
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 9223
;; LENGTH: 921
;; TYPE: PRT
;; ORGANISM: Drosophila
US-60-173-464-9223

Query Match 10.8%; Score 243.5; DB 27; Length 921;
Best Local Similarity 24.8%; Pred. No. 3.6e-10;

Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

QY 25 POIYAVKAPNTTIOFPANLQLPCTVLKSNGLML-----VSPQOTVTRAEYTS 76
DB 133 PPSPTITLSTLNTGTPA-----LVKTDNGFQLRVGTTGPTVQTITNTSNN 184
QY 77 NITSPAVPANPQVTKICTVPSNSQ-----LIKKAVTPYKLAQIGTVVTVTP 127
DB 185 NITSTTNHPTTQ-IRLQTVPAASMTNTATSNIIIVNSVASSGVANSOPPHLTQLANQ 243
QY 128 KPSSVQSAVPTSVTVTPGKPLNTVTTLKPSLSGASSTPSNEPMLKANSAAVQINLSP 187
DB 244 APQLPQITQITIPAOOSQOQOVNNVSSAGGATAVSSTTA-----ATT 287
QY 188 TMLENVK-KCKNELAMLIKACSSQSPDMGONVKLYEQLDAKIEAEFFTRKLYELK 246
DB 288 TQGGTKCKCKFLANLIEL--STREPKPVEKNVTLQELVNAVVEEFCDRLERLN 345
QY 247 SSPQHLVPLFKSVVALROL-----LPSQSFQIQ--- 276
DB 346 ASPQCLIGFLKSLPLRQALYTKELVIEGIKPPQHVLAGLSQQLPKIQAOIRPIG 405
QY 277 ---QCVQOTSSDMVIATCTTVTTSPPVYTTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGQTOVRMI--TPNALGTPRPRTIGHITTSKOPN--IRLPTAPRLVNTGIRT 460
QY 330 LNPLAGPVGAKAGVYTLHSGPTATGCTAGTGLQTSKPLVTSVANTVTVTSIQPEKP 389
DB 461 QIP-SLOVPGQANIVQIR--GPQHAQLQRTGSVOIRATTRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPNAVTFGETSGAALCLPSVKP 420
DB 507 -----KLTAVKVGQTOIKAI--TPSLHP 527

RESULT 15

US-60-173-464-20611
;; Sequence 20611, Application US/60173464
;; GENERAL INFORMATION:
;; APPLICANT: LI, Peter W.D.
;; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
;; TITLE OF INVENTION: THEROPO
;; FILE REFERENCE: CLO00173
;; CURRENT APPLICATION NUMBER: US/60/173,464
;; CURRENT FILING DATE: 1999-12-29
;; NUMBER OF SEQ ID NOS: 30269
;; SOFTWARE: FASTSEQ for Windows Version 4.0
;; SEQ ID NO 20611
;; LENGTH: 921
;; TYPE: PRT
;; ORGANISM: Drosophila
US-60-173-464-20611

Query Match 10.8%; Score 243.5; DB 27; Length 921;
Best Local Similarity 24.8%; Pred. No. 3.6e-10;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

QY 25 POIYAVKAPNTTIOFPANLQLPCTVLKSNGLML-----VSPQOTVTRAEYTS 76
DB 133 PPSPTITLSTLNTGTPA-----LVKTDNGFQLRVGTTGPTVQTITNTSNN 184
QY 77 NITSPAVPANPQVTKICTVPSNSQ-----LIKKAVTPYKLAQIGTVVTVTP 127
DB 185 NITSTTNHPTTQ-IRLQTVPAASMTNTATSNIIIVNSVASSGVANSOPPHLTQLANQ 243
QY 128 KPSSVQSAVPTSVTVTPGKPLNTVTTLKPSLSGASSTPSNEPMLKANSAAVQINLSP 187
DB 244 APQLPQITQITIPAOOSQOQOVNNVSSAGGATAVSSTTA-----ATT 287
QY 188 TMLENVK-KCKNELAMLIKACSSQSPDMGONVKLYEQLDAKIEAEFFTRKLYELK 246
DB 288 TQGGTKCKCKFLANLIEL--STREPKPVEKNVTLQELVNAVVEEFCDRLERLN 345
QY 247 SSPQHLVPLFKSVVALROL-----LPSQSFQIQ--- 276
DB 346 ASPQCLIGFLKSLPLRQALYTKELVIEGIKPPQHVLAGLSQQLPKIQAOIRPIG 405
QY 277 ---QCVQOTSSDMVIATCTTVTTSPPVYTTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGQTOVRMI--TPNALGTPRPRTIGHITTSKOPN--IRLPTAPRLVNTGIRT 460
QY 330 LNPLAGPVGAKAGVYTLHSGPTATGCTAGTGLQTSKPLVTSVANTVTVTSIQPEKP 389
DB 461 QIP-SLOVPGQANIVQIR--GPQHAQLQRTGSVOIRATTRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPNAVTFGETSGAALCLPSVKP 420
DB 507 -----KLTAVKVGQTOIKAI--TPSLHP 527

Search completed: February 16, 2003, 22:18:24
Job time : 165.507 secs


```

FILE REFERENCE: 0342/1G548-US1
CURRENT APPLICATION NUMBER: PCT/US01/20592A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592A-14

```

```

Query Match      22.8%; Score 513; DB 1; Length 1023;
Best Local Similarity 33.4%; Pred. No. 3,4e-32;
Matches 154; Conservative 62; Mismatches 127; Indels 118; Gaps 19;

```

```

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGGLM 61
DB 414 PRATSGIRATLPTVLAPRLPQ-----PQNPNTIQ--NFQLPQWLVRSNGQL 464
QY 62 LVSPQQTVTR-----AETSNITSRAVAVNPQVAVICVPPNSSQLIKKVAATPVKKL 115
DB 465 MI-PQALAQMOQAHAQPOPTMAPRPAPTPSAPPVQISTVQAGPTIAR-QVTP----- 518
QY 116 AQGTIVVTVTPKPSVQSAVAVPTSVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 519 -----TTTIKQV---SQAOITVQPSATIQRSPPQPOLVLGGAQTAISGTAATVQGTGPQ 571
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFLMLIKLACSSGQSPSEMGQNVKKYLEOL 227
DB 572 RTVPGATTTSSAATE-----TMENVKCKNFLSTLIKLASSGQSTETANVKELOVL 625
QY 228 LDKAIEEETFRKLYVELKSSPOPHLVPLFKKSVVALROLPLNSQSTLOOCVQO-----TS 283
DB 626 LDGKIEADEFTSRILKRLNSSPOPYLVPLFKRSILPALROLPLDPSAATIQSOQOQPPPTPS 685
QY 284 SDKVIATCTTVTTSPTVTTVSSQSEKSIIVSGATAPRTVSQVOTLNPLAGPVGAKAGV 343
DB 686 Q-----ATTALTAVALVSSVQRTAGKTAATVTSALQPVLSL----- 722
QY 344 VTLHVGPTAATGTTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAATVLSLPAV 403
DB 723 -----TQPTQVGVGKQGPPTPLVIO-----QPPKP-----GALIRPPQV 756
QY 404 TFGETSGAATCLPSVKPVVSVFCMDHICKPVIGTPVQIKLQ 444
DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQOVNLSE 785

```

RESULT 11
US-09-893-519A-14
Sequence 14, Application US/09893519A
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao

```

APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
RELEVANT RESIDUES: (1)..(1023)
US-09-893-519A-14

```

```

Query Match      22.8%; Score 513; DB 22; Length 1023;
Best Local Similarity 33.4%; Pred. No. 3,4e-32;
Matches 154; Conservative 62; Mismatches 127; Indels 118; Gaps 19;

```

```

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGGLM 61
DB 414 PRATSGIRATLPTVLAPRLPQ-----PQNPNTIQ--NFQLPQWLVRSNGQL 464
QY 62 LVSPQQTVTR-----AETSNITSRAVAVNPQVAVICVPPNSSQLIKKVAATPVKKL 115
DB 465 MI-PQALAQMOQAHAQPOPTMAPRPAPTPSAPPVQISTVQAGPTIAR-QVTP----- 518
QY 116 AQGTIVVTVTPKPSVQSAVAVPTSVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 519 -----TTTIKQV---SQAOITVQPSATIQRSPPQPOLVLGGAQTAISGTAATVQGTGPQ 571
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFLMLIKLACSSGQSPSEMGQNVKKYLEOL 227
DB 572 RTVPGATTTSSAATE-----TMENVKCKNFLSTLIKLASSGQSTETANVKELOVL 625
QY 228 LDKAIEEETFRKLYVELKSSPOPHLVPLFKKSVVALROLPLNSQSTLOOCVQO-----TS 283
DB 626 LDGKIEADEFTSRILKRLNSSPOPYLVPLFKRSILPALROLPLDPSAATIQSOQOQPPPTPS 685
QY 284 SDKVIATCTTVTTSPTVTTVSSQSEKSIIVSGATAPRTVSQVOTLNPLAGPVGAKAGV 343
DB 686 Q-----ATTALTAVALVSSVQRTAGKTAATVTSALQPVLSL----- 722
QY 344 VTLHVGPTAATGTTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAATVLSLPAV 403
DB 723 -----TQPTQVGVGKQGPPTPLVIO-----QPPKP-----GALIRPPQV 756
QY 404 TFGETSGAATCLPSVKPVVSVFCMDHICKPVIGTPVQIKLQ 444
DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQOVNLSE 785

```

RESULT 12
PCT-US02-30474-3271
Sequence 3271, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aidong J.
APPLICANT: Zhao, Qiang A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Weinman, Tom

PRIOR APPLICATION NUMBER: PCT/US01/14827
PRIOR FILING DATE: 2001-05-16
NUMBER OF SEQ ID NOS: 3334
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO: 1665
LENGTH: 1051
TYPE: PRT
ORGANISM: Homo sapiens
US-60-324-631-1665

Query Match 23.2%; Score 521; DB 27; Length 1051;
Best Local Similarity 34.0%; Pred. No. 7,86-33;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

OY 13 PKVSSG-----PRLPAQIVAAKAPNTTITOPANQLPPTGYLIKNSGPILM 61
DB 382 PRATTSIGIRATLPTVLARLP-----PQNPTNIQ---NFQLPPGAVLVNSENGL 432
OY 62 LVSPOQVTR-----AETTSNITSRAVPANPQTVKICTVPNSSQLIKKVAVTPVKKL 115
DB 433 MI-POALAQMOQAHAQOTTMARPAPTSPAPVOISTVQAPGPIIAR-QVTP----- 486
OY 116 AIGTIVTVTPKPSVQSAVPTSVYVTPGKPLNTV--TLKPSLGLASS-----TPS 168
DB 487 ----TTLIKOV---SOAQTVQPSATLORSPEGVQPOLVGAAGTASLGTATAVQGTPO 539
OY 169 NE-PNLKENSAAVOINLSPTMLENVKCKNFAMLIKLAGSGSGSPENGQNVKLEVL 227
DB 540 RTVPAGTTSSATE-----TMENVKCKNFSTLIKLAGSGSGKOSTETRAANKVLYOL 593
OY 228 LDKAIEAEFTKRLVELKSSPQPHLVPLKKSVALROLPLNSGFTIOCCVQO---TS 283
DB 594 LDKAIEAEFTKRLVELKSSPQPHLVPLKKSVALROLPLNSGFTIOCCVQO---TS 283
OY 284 SDMVATCTTIVTTSVYVTTVSSGSEKSIIVSGATFARTVSVOVLNLNLAGVGAAGV 343
DB 654 Q-----ATTALTAVLVSSVORTAGKTAATVTSALQPPVLSL----- 690
OY 344 VTLHSVPTAAGTGTAGLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLP 403
DB 691 -----TQPTQGVGKQGGQPTPLVIO-----QPKP-----GALIRPPV 724
OY 404 TGEISGAALCLPSVKPVVSPFCMDHICKPVIGTPOIKL--AQPGVYL 449
DB 725 TLTQT-----PMVALRQPH-NRIMLTTPQOIQLNPLQPVVV 760

RESULT 9
PCT-US01-20592-14
Sequence 14, Application PC/TUS0120592
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTI-FUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/16548-US1
CURRENT APPLICATION NUMBER: PCT/US01/20592
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29

PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: human genbank accession #: CAA72189
NAME/KEY: misc.feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592-14

Query Match 22.8%; Score 513; DB 1; Length 1023;
Best Local Similarity 33.4%; Pred. No. 3,46-32;
Matches 154; Conservative 62; Mismatches 127; Indels 118; Gaps 19;

OY 13 PKVSSG-----PRLPAQIVAAKAPNTTITOPANQLPPTGYLIKNSGPILM 61
DB 414 PRATTSIGIRATLPTVLARLP-----PQNPTNIQ---NFQLPPGAVLVNSENGL 464
OY 62 LVSPOQVTR-----AETTSNITSRAVPANPQTVKICTVPNSSQLIKKVAVTPVKKL 115
DB 465 MI-POALAQMOQAHAQOTTMARPAPTSPAPVOISTVQAPGPIIAR-QVTP----- 518
OY 116 AIGTIVTVTPKPSVQSAVPTSVYVTPGKPLNTV--TLKPSLGLASS-----TPS 168
DB 519 ----TTLIKOV---SOAQTVQPSATLORSPEGVQPOLVGAAGTASLGTATAVQGTPO 571
OY 169 NE-PNLKENSAAVOINLSPTMLENVKCKNFAMLIKLAGSGSGSPENGQNVKLEVL 227
DB 572 RTVPAGTTSSATE-----TMENVKCKNFSTLIKLAGSGSGKOSTETRAANKVLYOL 625
OY 228 LDKAIEAEFTKRLVELKSSPQPHLVPLKKSVALROLPLNSGFTIOCCVQO---TS 283
DB 626 LDKAIEAEFTKRLVELKSSPQPHLVPLKKSVALROLPLNSGFTIOCCVQO---TS 283
OY 284 SDMVATCTTIVTTSVYVTTVSSGSEKSIIVSGATFARTVSVOVLNLNLAGVGAAGV 343
DB 686 Q-----ATTALTAVLVSSVORTAGKTAATVTSALQPPVLSL----- 722
OY 344 VTLHSVPTAAGTGTAGLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLP 403
DB 723 -----TQPTQGVGKQGGQPTPLVIO-----QPKP-----GALIRPPV 756
OY 404 TGEISGAALCLPSVKPVVSPFCMDHICKPVIGTPOIKL 444
DB 757 TLTQT-----PMVALRQPH-NRIMLTTPQOVNLSE 785

RESULT 10
PCT-US01-20592A-14
Sequence 14, Application PC/TUS0120592A
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTI-FUNGAL COMPOUNDS AND METHODS OF USE

```

: APPLICANT: Ghosh, Malabika
: APPLICANT: Wang, Dunrul
: APPLICANT: Ma, Yungqing
: APPLICANT: Asundi, Vinod
: APPLICANT: Wang, Zhiwei
: APPLICANT: Meng, Gezhi
: APPLICANT: Haley-Vicente, Dana
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: FILE OF INVENTION: Polypeptides
: FILE REFERENCE: 810CIP PCT
: CURRENT APPLICATION NUMBER: PCT/US02/30474
: CURRENT FILING DATE: 2002-09-24
: PRIOR APPLICATION NUMBER: US 60/324,631
: PRIOR FILING DATE: 2001-09-24
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03800
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 09/515,126
: PRIOR FILING DATE: 2000-02-28
: Remaining Prior Application data removed - See file wrapper or PALM.
: NUMBER OF SEQ ID NOS: 3476
: SOFTWARE: PL_FL-genes Version 6.0
: SEQ ID NO 1660
: LENGTH: 1051
: TYPE: PRT
: ORGANISM: Homo sapiens
: PCT-US02-30474-1660

Query Match          23.2% Score 521; DB 1; Length 1051;
Best Local Similarity 34.0%; Pred. No. 7.8e-33;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTIQFPANQLPPTVLKNSGFLM 61
DB 382 PTLATTSIGRATLPTVLAPRLPQP-----PQNTNIO--NFQDLPQMTLVRENGQLL 432
QY 62 LVSPQQTIVTR-----AETTSNITSRPAVPANPQTVKICIVPNSSQLIKKVAVTPVAKL 115
DB 433 MI-PQQAIAQMOQAQAHAPQTTMAPRPATPTAPRPVQISTVQACGTPRIAR-QVTP----- 486
QY 116 AAGCTTAVTVTPKRSVQSAVAVPSTVYTPGKRLNTYT--TLKPPSSIGASS-----TPS 168
DB 487 -----TTTIKQV--SQAQTVYQPSATLQSRSPGVOPQLVLGGAATASGATTAVQGTGP 539
QY 169 NE-PNLKAENSAVAQVIMLSPTMLENVKCKKNFLMLIKLACSGSQSPMGONVKVLEQL 227
DB 540 RTVPGATTTSSAATE-----TMENVKCKKNFLSTLIKILKSSGSGQSTETANVKEIYNL 593
QY 228 LDKAIEAEETFRKLYVELKSSPOPHLVLPFLKKSVALRQLLPPNSQSFIOQCVQO-----TS 283
DB 594 LDKIEAEDEFTSRILYRELNSPQPYLVLPFLKRSIPALROLPPDSAAFIOGSOQOPPPPTS 653
QY 284 SDVVIATCTTAVTSPVTTVSSQSEKSIIVSGATAPRVSQVTLNPLAGPVGAKAGV 343
DB 654 Q-----ATTALTAVALVSSVQFTAGKTATVTSALQPVLSL-----690
QY 344 VTLHSVGPATAGCTGTAGTGLQTSKPLVNTSVANTVTTVSLQPKPVSVGTAVTLSLPAV 403
DB 691 -----TQPTQVGVKGKQSGPTPLVIO-----QPKP-----GALIRPPQV 724
```

```

QY 404 TFGENGAITCLPSVKPVYSCWMDHICKPVIGTPVQIKL--AQGCPVL 449
DB 725 TLTQF-----PMVALRQPH-NRIMMLTTPQOIQILNPLQPVV 760

RESULT 8
US-60-324-631-1665
: Sequence 1665, Application US/60324631
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Zhang, Jie
: APPLICANT: Ren, Feiyun
: APPLICANT: Xue, Aiding J.
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Wehrman, Tom
: APPLICANT: Zhou, Ping
: APPLICANT: Ghosh, Malabika
: APPLICANT: Wang, Dunrul
: APPLICANT: Ma, Yungqing
: APPLICANT: Asundi, Vinod
: APPLICANT: Wang, Zhiwei
: APPLICANT: Meng, Gezhi
: APPLICANT: Haley-Vicente, Dana
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: FILE OF INVENTION: Polypeptides
: FILE REFERENCE: 810
: CURRENT APPLICATION NUMBER: US/60/324,631
: CURRENT FILING DATE: 2001-09-24
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03800
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 09/515,126
: PRIOR FILING DATE: 2000-02-28
: PRIOR APPLICATION NUMBER: US 09/577,409
: PRIOR FILING DATE: 2000-05-18
: PRIOR APPLICATION NUMBER: PCT/US01/04927
: PRIOR FILING DATE: 2001-02-26
: PRIOR APPLICATION NUMBER: US 09/519,705
: PRIOR FILING DATE: 2000-03-07
: PRIOR APPLICATION NUMBER: US 09/574,454
: PRIOR FILING DATE: 2000-05-19
: PRIOR APPLICATION NUMBER: PCT/US01/04941
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: US 09/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: US 09/649,167
: PRIOR FILING DATE: 2000-08-23
: PRIOR APPLICATION NUMBER: PCT/US01/08631
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: US 09/552,929
: PRIOR FILING DATE: 2000-04-18
: PRIOR APPLICATION NUMBER: US 09/770,160
: PRIOR FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: PCT/US01/08656
: PRIOR FILING DATE: 2001-04-18
: PRIOR APPLICATION NUMBER: US 09/577,408
: PRIOR FILING DATE: 2000-05-18
```

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
FILE REFERENCE: CLO00929
CURRENT APPLICATION NUMBER: US/60/243,468
CURRENT FILING DATE: 2000-10-27
NUMBER OF SEQ ID NOS: 2121
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1271
LENGTH: 685
TYPE: PRT
ORGANISM: HUMAN
US-60-243-468-1271

Query Match 88.3%; Score 1986; DB 27; Length 685;
Best Local Similarity 90.5%; Pred. No. 3,2e-153;
Matches 409; Conservative 0; Mismatches 7; Indels 36; Gaps 1;

QY 1 GTLVKRVAVSAPKVSQGPRLPAQIVAVKAPNTTIOFPANLQRPCTVLKSSGGL 60
DB 67 GTLVKRVAVSAPKVSQGPRLPAQIVAVKAPNTTIOFPANLQRPCTVLKSSGGL 126
QY 61 MLVSPQGVTRATTSNITSRPAVPANPQTVKICTVPNSSQLIKKVAATPVKLAQIGT 120
DB 127 MLVSPQGVTRATTSNITSRPAVPANPQTVKICTVPNSSQLIKKVAATPVKLAQIGT 163
QY 121 TVVTVPKRSSVQSAVAVPVSVTVPKPLNTVTTLKPSLGAASSPSPNEPNKAKNSAA 180
DB 164 -----SVAVPSTSVTVTPKPLNTVTTLKPSLGAASSPSPNEPNKAKNSAA 210
QY 181 VOINISPTMLENVKCKKNFLAMLIKACSGSPKPGONVKIIEVDLAKTEAEFTTK 240
DB 211 VOINISPTMLENVKCKKNFLAMLIKACSGSPKPGONVKIIEVDLAKTEAEFTTK 270
QY 241 LVELKSSPQHLVPLKRSVVALROLPLNSOSFIOQVQOTSDDVIACTTTVTTSY 300
DB 271 LVELKSSPQHLVPLKRSVVALROLPLNSOSFIOQVQOTSDDVIACTTTVTTSY 330
QY 301 VTTVSSSSSEKSIIVSGATPRTVSVOQLNPLAGFVGAKAGVTLHSVGPAAATGTTA 360
DB 331 VTTVSSSSSEKSIIVSGATPRTVSVOQLNPLAGFVGAKAGVTLHSVGPAAATGTTA 390
QY 361 GGLIOTSPLVTSVANVTTVSLOPEKPVSGTAVTSLPATEFETGGAALCLPSVAP 420
DB 391 GGLIOTSPLVTSVANVTTVSLOPEKPVSGTAVTSLPATEFETGGAALCLPSVAP 450
QY 421 VVSCMDHICKPVIGTPVOIKLAQPGVLSOP 452
DB 451 VVSCMDHICKPVIGTPVOIKLAQPGVLSOP 482

RESULT 6
PCT-US94-01114-16
Sequence 16, Application PC/TUS9401114
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comal, Lucio
APPLICANT: Dylact, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESS: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01114
FILING DATE: 28-JAN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: FP-57650-2/AUT/RAO
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01114-16

Query Match 23.2%; Score 521; DB 1; Length 737;
Best Local Similarity 34.0%; Pred. No. 4,5e-33;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PPKVSSG-----FRAPQIVAVKAPNTTIOFPANLQRPCTVLKSSGGL 61
DB 68 PPKVSSG-----FRAPQIVAVKAPNTTIOFPANLQRPCTVLKSSGGL 118
QY 62 LVSPQGVTR-----AETTSNITSRPAVPANPQTVKICTVPNSSQLIKKVAATPVKTL 115
DB 119 LVSPQGVTR-----AETTSNITSRPAVPANPQTVKICTVPNSSQLIKKVAATPVKTL 172
QY 116 AQIGTVTVTPKRSVQSAVAVPVSVTVPKPLNTVTTLKPSLGAASSPSPNEPNKAKNSAA 168
DB 173 AQIGTVTVTPKRSVQSAVAVPVSVTVPKPLNTVTTLKPSLGAASSPSPNEPNKAKNSAA 225
QY 169 NE-PNKAENSAVQINLSPTMLENVKCKKNFLAMLIKACSGSPKPGONVKIIEVDLAKTEAEFTTK 227
DB 226 NE-PNKAENSAVQINLSPTMLENVKCKKNFLAMLIKACSGSPKPGONVKIIEVDLAKTEAEFTTK 279
QY 228 LPAKIEAEFTTKLYELKSSPQHLVPLKRSVVALROLPLNSOSFIOQVQOTSDDVIACTTTVTTSY 283
DB 280 LPAKIEAEFTTKLYELKSSPQHLVPLKRSVVALROLPLNSOSFIOQVQOTSDDVIACTTTVTTSY 339
QY 284 SDVIAVTCCTTVTSVAVPVSVTVPKPLNTVTTLKPSLGAASSPSPNEPNKAKNSAA 343
DB 340 SDVIAVTCCTTVTSVAVPVSVTVPKPLNTVTTLKPSLGAASSPSPNEPNKAKNSAA 376
QY 344 VILHSVGPAAATGTTAGGLIOTSPLVTSVANVTTVSLOPEKPVSGTAVTSLPATEFETGGAALCLPSVAP 403
DB 377 VILHSVGPAAATGTTAGGLIOTSPLVTSVANVTTVSLOPEKPVSGTAVTSLPATEFETGGAALCLPSVAP 410
QY 404 TFEETSGAALCLPSVAVPVSFCMDHICKPVIGTPVOIKL--AOPGVL 449
DB 411 TFEETSGAALCLPSVAVPVSFCMDHICKPVIGTPVOIKL--AOPGVL 446

RESULT 7
PCT-US02-30474-1660
Sequence 1660, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Xue, Aidong J.
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping

```

: PRIOR APPLICATION NUMBER: US 60/328,185
: PRIOR FILING DATE: 2001-10-05
: NUMBER OF SEQ ID NOS: 66
: SOFTWARE: PERL Program
: SEQ ID NO 25
: LENGTH: 865
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: Incyte ID No: 4398735CD1
PCT-US02-25829-25

Query Match      99.5%  Score 2237.5; DB 1; Length 865;
Best Local Similarity 99.3%; Pred. No. 1.1e-173;
Matches 452; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 GTTATKAPVAPPKVSSGRLPAPQIVAKAPMTTITOPANQLP---GTVLINSNS 57
Db 67 GTTATKAPVAPPKVSSGRLPAPQIVAKAPMTTITOPANQLP---GTVLINSNS 126
QY 58 GPLMVSPOQTVTRAETTSNITSRPAPVAPNDQTVKICTVNSSQLIKKVAVPPVKKLAQ 117
Db 127 GPLMVSPOQTVTRAETTSNITSRPAPVAPNDQTVKICTVNSSQLIKKVAVPPVKKLAQ 186
QY 118 IGTTVTVTVKRPSSVQAVPTSVYVTPGKPLNTVTTLKPSLGASSTPSNEPNLKAEN 177
Db 187 IGTTVTVTVKRPSSVQAVPTSVYVTPGKPLNTVTTLKPSLGASSTPSNEPNLKAEN 246
QY 178 SAAVQINLSPMLLENVKKCNFLMLIKLACSGSOSPEMGONVKKLEOLDIAIEAEF 237
Db 247 SAAVQINLSPMLLENVKKCNFLMLIKLACSGSOSPEMGONVKKLEOLDIAIEAEF 306
QY 238 TRRLVELKSSPOPHLVPLFKSVVALROLPLNSQSFIOQCVOQTSSDMVYATCTTVT 297
Db 307 TRRLVELKSSPOPHLVPLFKSVVALROLPLNSQSFIOQCVOQTSSDMVYATCTTVT 366
QY 298 SPVTTTSSQSEKSTIVSGATAPRTVSVQTLNPLAGPVGAKGVYTLHSVGTAAATGC 357
Db 367 SPVTTTSSQSEKSTIVSGATAPRTVSVQTLNPLAGPVGAKGVYTLHSVGTAAATGC 426
QY 358 TTAGTGLQTSKPLVTSVANTVTTVSLOPEKPVVSGTAVTLSPVAFGEFGSAAICLPS 417
Db 427 TTAGTGLQTSKPLVTSVANTVTTVSLOPEKPVVSGTAVTLSPVAFGEFGSAAICLPS 486
QY 418 VKPVSEFCMDHICKPVIGTPVQIKLAQPGVLSQP 452
Db 487 VKPVSEFCMDHICKPVIGTPVQIKLAQPGVLSQP 521

RESULT 4
PCT-US02-29964-410
: Sequence 410, Application PC/TUS0229964
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Ren, Feiyang
: APPLICANT: Zhang, Jie
: APPLICANT: Zhao, Qing A.
: APPLICANT: Wang, Jian-Rui
: APPLICANT: Xue, Aifeng J.
: APPLICANT: Wang, Dunrui
: APPLICANT: Ghosh, Malabika
: APPLICANT: Asundi, Vinod
: APPLICANT: Wehrman, Tom
: APPLICANT: Zhou, Ping
: APPLICANT: Goodrich, Ryle W.
: APPLICANT: Meng, Gezhil
: APPLICANT: Haley-Vicente, Dana
: APPLICANT: Drmanac, Radoje T
: TITLE OF INVENTION: Novel Nucleic Acids and
: FILE REFERENCE: 809ACIP PCT
: CURRENT APPLICATION NUMBER: PCT/US02/29964

```

```

: CURRENT FILING DATE: 2002-09-19
: PRIOR APPLICATION NUMBER: US 60/323,739
: PRIOR FILING DATE: 2001-09-19
: PRIOR APPLICATION NUMBER: US 09/488,725
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 09/552,317
: PRIOR FILING DATE: 2000-04-25
: PRIOR APPLICATION NUMBER: PCT/US00/35017
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: US 09/491,404
: PRIOR FILING DATE: 2000-01-25
: PRIOR APPLICATION NUMBER: PCT/US01/02623
: PRIOR FILING DATE: 2001-01-25
: PRIOR APPLICATION NUMBER: US 09/496,914
: PRIOR FILING DATE: 2000-02-03
: PRIOR APPLICATION NUMBER: US 09/560,875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: PCT/US01/03800
: PRIOR FILING DATE: 2001-02-05
: PRIOR APPLICATION NUMBER: US 09/515,126
: Remaining Prior Application data removed - See File Wrapper or PALM.
: SOFTWARE: PL_FL_genes Version 6.0
: SEQ ID NO 410
: LENGTH: 843
: TYPE: PRT
: ORGANISM: Homo sapiens
PCT-US02-29964-410

Query Match      97.1%  Score 2184; DB 1; Length 843;
Best Local Similarity 96.5%; Pred. No. 2.6e-169;
Matches 445; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GTTATKAPVAPPKVSSGRLPAPQIVAKAPMTTITOPANQLP---GTVLINSNS 60
Db 43 GTTATKAPVAPPKVSSGRLPAPQIVAKAPMTTITOPANQLP---GTVLINSNS 102
QY 61 MLVSPQOQTVTRAETTSNITSRPAPVAPNDQTVKICTVNSSQLIKKVAVPPVKKLAQ 120
Db 103 MLVSPQOQTVTRAETTSNITSRPAPVAPNDQTVKICTVNSSQLIKKVAVPPVKKLAQ 162
QY 121 TVTTVTVKRPSSVQAVPTSVYVTPGKPLNTVTTLKPSLGASSTPSNEPNLKAEN 180
Db 163 TVTTVTVKRPSSVQAVPTSVYVTPGKPLNTVTTLKPSLGASSTPSNEPNLKAEN 222
QY 181 VOINLSPMLLENVKKCNFLMLIKLACSGSOSPEMGONVKKLEOLDIAIEAEFTRK 240
Db 223 VOINLSPMLLENVKKCNFLMLIKLACSGSOSPEMGONVKKLEOLDIAIEAEFTRK 282
QY 241 LVVELKSSPOPHLVPLFKSVVALROLPLNSQSFIOQCVOQTSSDMVYATCTTVTSPV 300
Db 283 LVVELKSSPOPHLVPLFKSVVALROLPLNSQSFIOQCVOQTSSDMVYATCTTVTSPV 342
QY 301 VTTTVSSQSEKSTIVSGATAPRTVSVQTLNPLAGPVGAKGVYTLHSVGTAAATGTTA 360
Db 343 VTTTVSSQSEKSTIVSGATAPRTVSVQTLNPLAGPVGAKGVYTLHSVGTAAATGTTA 402
QY 361 GTGLQTSKPLVTSVANTVTTVSLOPEKPVVSGTAVTLSPVAFGEFGSAAICLPSKV 420
Db 403 GTGLQTSKPLVTSVANTVTTVSLOPEKPVVSGTAVTLSPVAFGEFGSAAICLPSKV 462
QY 421 VVSEFCMDHICKPVIGTPVQIKLAQPGVLSQP 452
Db 463 VVSEFCMDHICKPVIGTPVQIKLAQPGVLSQP 494

RESULT 5
US-60-243-468-1271
: Sequence 1271, Application US/60243468
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,

```

```
OY 61 MLVSPQQTVAETTSNTSRPAVAPNPQTVKICTVPSNSSOLIKKVAATPVKKAQIGT 120
DB 61 MLVSPQQTVAETTSNTSRPAVAPNPQTVKICTVPSNSSOLIKKVAATPVKKAQIGT 120
OY 121 TVTTPVPSSVQSAVAPVTPVPGKPLNTVTTLKPSLCAASSPNEPMLKAENSA 180
DB 121 TVTTPVPSSVQSAVAPVTPVPGKPLNTVTTLKPSLCAASSPNEPMLKAENSA 180
OY 181 VOINISPTMLENVKCKNFMLIKLACSGSOPMEGONVKLVROLDAKTEAEFTFRK 240
DB 181 VOINISPTMLENVKCKNFMLIKLACSGSOPMEGONVKLVROLDAKTEAEFTFRK 240
OY 241 LVELEKSSPQPLVLPFLKSSVVALROLLPNSQSFIOQCVOQTSSDMVATATCTTVTTSFV 300
DB 241 LVELEKSSPQPLVLPFLKSSVVALROLLPNSQSFIOQCVOQTSSDMVATATCTTVTTSFV 300
OY 301 VTTTSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTLHSVGPAAAGGTTA 360
DB 301 VTTTSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTLHSVGPAAAGGTTA 360
OY 361 GTGLQTSKPLVTSVANTVTYVSLQPEKPVSGTAVTSLPAVTGEGTSGAICLPSPVP 420
DB 361 GTGLQTSKPLVTSVANTVTYVSLQPEKPVSGTAVTSLPAVTGEGTSGAICLPSPVP 420
OY 421 VVSEFMDHICKRPVIGTPVQIKLAOPGPLYLSP 452
DB 421 VVSEFMDHICKRPVIGTPVQIKLAOPGPLYLSP 452
```

RESULT 2

US-09-763-909-2

Sequence 2, Application US/09763909

GENERAL INFORMATION:

APPLICANT: Dikstein, Rivka

TITLE OF INVENTION: A TRANSCRIPTION FACTOR TRIID SUBUNIT,

TITLE OF INVENTION: TAF1105, POLYPEPTIDES, DNA ENCODING THEREFOR AND

FILE REFERENCE: 13005/002001

CURRENT APPLICATION NUMBER: US/09/763, 909

CURRENT FILING DATE: 2001-02-26

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 852

TYPE: PRT

ORGANISM: Homo sapiens

US-09-763-909-2

Query Match 100.0%; Score 2249; DB 21; Length 852;

Best Local Similarity 100.0%; Pred. No. 1.2e-174;

Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 GTLVTKVAPVAPKPVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPQGTIVIKNSGPL 60
DB 1 GTLVTKVAPVAPKPVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPQGTIVIKNSGPL 60
OY 61 MLVSPQQTVAETTSNTSRPAVAPNPQTVKICTVPSNSSOLIKKVAATPVKKAQIGT 120
DB 61 MLVSPQQTVAETTSNTSRPAVAPNPQTVKICTVPSNSSOLIKKVAATPVKKAQIGT 120
OY 121 TVTTPVPSSVQSAVAPVTPVPGKPLNTVTTLKPSLCAASSPNEPMLKAENSA 180
DB 121 TVTTPVPSSVQSAVAPVTPVPGKPLNTVTTLKPSLCAASSPNEPMLKAENSA 180
OY 181 VOINISPTMLENVKCKNFMLIKLACSGSOPMEGONVKLVROLDAKTEAEFTFRK 240
DB 181 VOINISPTMLENVKCKNFMLIKLACSGSOPMEGONVKLVROLDAKTEAEFTFRK 240
OY 241 LVELEKSSPQPLVLPFLKSSVVALROLLPNSQSFIOQCVOQTSSDMVATATCTTVTTSFV 300
DB 241 LVELEKSSPQPLVLPFLKSSVVALROLLPNSQSFIOQCVOQTSSDMVATATCTTVTTSFV 300
```

```
OY 301 VTTTSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTLHSVGPAAAGGTTA 360
DB 301 VTTTSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTLHSVGPAAAGGTTA 360
OY 361 GTGLQTSKPLVTSVANTVTYVSLQPEKPVSGTAVTSLPAVTGEGTSGAICLPSPVP 420
DB 361 GTGLQTSKPLVTSVANTVTYVSLQPEKPVSGTAVTSLPAVTGEGTSGAICLPSPVP 420
OY 421 VVSEFMDHICKRPVIGTPVQIKLAOPGPLYLSP 452
DB 421 VVSEFMDHICKRPVIGTPVQIKLAOPGPLYLSP 452
```

RESULT 3

PCT-US02-25829-25

Sequence 25, Application PC/TUS0225829

GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.

APPLICANT: BAROSSO, Ines

APPLICANT: BAUGHN, Mariah R.

APPLICANT: BECHA, Shanya D.

APPLICANT: BLAKE, Julie J.

APPLICANT: BOROWSKY, Mark L.

APPLICANT: BURFORD, Neil

APPLICANT: DUGGAN, Brendan M.

APPLICANT: ELLIOTT, Vicki S.

APPLICANT: EMERLING, Brooke M.

APPLICANT: FORSYTHE, Ian J.

APPLICANT: GIERZEN, Kimberly J.

APPLICANT: GORVAD, Ann E.

APPLICANT: GRIFFIN, Jennifer A.

APPLICANT: HAFALIA, April J.A.

APPLICANT: HONCHELL, Cynthia D.

APPLICANT: ISON, Craig H.

APPLICANT: KHAN, Farrah A.

APPLICANT: LAU, Preeti G.

APPLICANT: LEE, Ernestine A.

APPLICANT: LEE, Sally

APPLICANT: LEE, Soo Yeun

APPLICANT: LI, Joana X.

APPLICANT: LU, Dzung Alina M.

APPLICANT: LEHR-MASON, Patricia M.

APPLICANT: NGUYEN, Danielle B.

APPLICANT: RAMKUMAR, Jayalaxmi

APPLICANT: SPRAGUE, William W.

APPLICANT: TANG, Y. Tom

APPLICANT: THANGAVELOU, Kavitha

APPLICANT: THORNTON, Michael

APPLICANT: TRAN, Uyen K.

APPLICANT: WALIA, Narinder K.

APPLICANT: WARREN, Bridget A.

APPLICANT: XU, Yuning

APPLICANT: YAO, Monique G.

APPLICANT: YUE, Henry

APPLICANT: YUE, Huibin

APPLICANT: ZEBARADIAN, Yeganeh

TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS

FILE REFERENCE: PF-1146 PCT

CURRENT APPLICATION NUMBER: PCT/US02/25829

CURRENT FILING DATE: 2002-08-14

PRIOR APPLICATION NUMBER: US 60/313,111

PRIOR FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: US 60/314,682

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/314,756

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/315,105

PRIOR FILING DATE: 2001-08-27

PRIOR APPLICATION NUMBER: US 60/316,751

PRIOR FILING DATE: 2001-08-31

PRIOR APPLICATION NUMBER: US 60/316,856

PRIOR FILING DATE: 2001-08-31

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 159.507 Seconds

(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452

Perfect score: 2249
Sequence: 1 GTLVTKVAPVSAAPKVSNGP.....VIGTPVQIKLAQPGVLSQP 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2249	100.0	801	1	PCT-US01-08631-39827
2	2249	100.0	852	21	US-09-763-909-2
3	2237.5	99.5	865	1	PCT-US02-25829-25
4	2184	97.1	843	1	PCT-US02-29964-410
5	1986	88.3	685	27	US-60-243-468-1271
6	521	23.2	737	1	PCT-US94-01114-16

7	521	23.2	1051	1	PCT-US02-30474-1660	Sequence 1660, Ap
8	521	23.2 <th>1051</th> <td>27</td> <td>US-60-324-631-1665</td> <td>Sequence 1665, Ap</td>	1051	27	US-60-324-631-1665	Sequence 1665, Ap
9	513	22.8 <th>1051</th> <td>1</td> <td>PCT-US01-20592-14</td> <td>Sequence 14, Appl</td>	1051	1	PCT-US01-20592-14	Sequence 14, Appl
10	513	22.8 <th>1051</th> <td>1</td> <td>PCT-US01-20592A-14</td> <td>Sequence 14, Appl</td>	1051	1	PCT-US01-20592A-14	Sequence 14, Appl
11	513	22.8 <th>1051</th> <td>22</td> <td>US-09-893-519A-14</td> <td>Sequence 14, Appl</td>	1051	22	US-09-893-519A-14	Sequence 14, Appl
12	509	22.6 <th>1083</th>	1083	1	PCT-US02-30474-3271	Sequence 3271, Ap
13	243.5	10.8 <td>899</td> <td>27</td> <td>US-60-167-217-11401</td> <td>Sequence 11401, A</td>	899	27	US-60-167-217-11401	Sequence 11401, A
14	243.5	10.8 <td>921</td> <td>27</td> <td>US-60-173-466-9223</td> <td>Sequence 9223, Ap</td>	921	27	US-60-173-466-9223	Sequence 9223, Ap
15	243.5	10.8 <td>921</td> <td>27</td> <td>US-60-173-466-9223</td> <td>Sequence 9223, Ap</td>	921	27	US-60-173-466-9223	Sequence 9223, Ap
16	241.5	10.7 <td>921</td> <td>1</td> <td>PCT-US94-01114-2</td> <td>Sequence 20611, A</td>	921	1	PCT-US94-01114-2	Sequence 20611, A
17	241.5	10.7 <td>921</td> <td>4</td> <td>US-08-013-412-2</td> <td>Sequence 2, Appl</td>	921	4	US-08-013-412-2	Sequence 2, Appl
18	241.5	10.7 <td>921</td> <td>4</td> <td>US-09-614-150-11376</td> <td>Sequence 11376, A</td>	921	4	US-09-614-150-11376	Sequence 11376, A
19	241.5	10.7 <td>921</td> <td>20</td> <td>US-09-614-150-24957</td> <td>Sequence 24957, A</td>	921	20	US-09-614-150-24957	Sequence 24957, A
20	241.5	10.7 <td>921</td> <td>27</td> <td>US-60-191-637-11408</td> <td>Sequence 11408, A</td>	921	27	US-60-191-637-11408	Sequence 11408, A
21	241.5	10.7 <td>921</td> <td>27</td> <td>US-60-191-637-25058</td> <td>Sequence 25058, A</td>	921	27	US-60-191-637-25058	Sequence 25058, A
22	241.5	10.7 <td>921</td> <td>27</td> <td>US-60-191-681-8933</td> <td>Sequence 8933, Ap</td>	921	27	US-60-191-681-8933	Sequence 8933, Ap
23	241.5	10.7 <td>921</td> <td>27</td> <td>US-60-191-681-19739</td> <td>Sequence 19739, A</td>	921	27	US-60-191-681-19739	Sequence 19739, A
24	219	9.7 <td>1953</td> <td>27</td> <td>US-60-185-361-566</td> <td>Sequence 566, App</td>	1953	27	US-60-185-361-566	Sequence 566, App
25	219	9.7 <td>2947</td> <td>27</td> <td>US-60-185-361-494</td> <td>Sequence 494, App</td>	2947	27	US-60-185-361-494	Sequence 494, App
26	219	9.7 <td>2947</td> <td>27</td> <td>US-60-185-361-938</td> <td>Sequence 938, Appl</td>	2947	27	US-60-185-361-938	Sequence 938, Appl
27	216.5	9.6 <td>1149</td> <td>27</td> <td>US-60-185-361-565</td> <td>Sequence 565, App</td>	1149	27	US-60-185-361-565	Sequence 565, App
28	209.5	9.3	528	22	US-09-840-746-20	Sequence 36210, A
29	202	9.0 <td>1795</td> <td>20</td> <td>US-09-614-150-35824</td> <td>Sequence 35824, A</td>	1795	20	US-09-614-150-35824	Sequence 35824, A
30	202	9.0 <td>1795</td> <td>27</td> <td>US-60-191-637-35824</td> <td>Sequence 35824, A</td>	1795	27	US-60-191-637-35824	Sequence 35824, A
31	202	9.0 <td>1795</td> <td>27</td> <td>US-60-191-681-28036</td> <td>Sequence 28036, A</td>	1795	27	US-60-191-681-28036	Sequence 28036, A
32	201.5	9.0 <td>864</td> <td>27</td> <td>US-60-173-386-314</td> <td>Sequence 314, App</td>	864	27	US-60-173-386-314	Sequence 314, App
33	201.5	9.0 <td>864</td> <td>27</td> <td>US-60-173-386-314</td> <td>Sequence 314, App</td>	864	27	US-60-173-386-314	Sequence 314, App
34	201.5	9.0 <td>864</td> <td>27</td> <td>US-60-173-386-314</td> <td>Sequence 314, App</td>	864	27	US-60-173-386-314	Sequence 314, App
35	200.5	8.9 <td>1367</td> <td>18</td> <td>US-09-487-558-108</td> <td>Sequence 324, App</td>	1367	18	US-09-487-558-108	Sequence 324, App
36	200.5	8.9 <td>1367</td> <td>18</td> <td>US-09-487-558-108</td> <td>Sequence 108, App</td>	1367	18	US-09-487-558-108	Sequence 108, App
37	200.5	8.9 <td>1367</td> <td>22</td> <td>US-09-801-366-108</td> <td>Sequence 108, App</td>	1367	22	US-09-801-366-108	Sequence 108, App
38	200.5	8.9 <td>1367</td> <td>22</td> <td>US-09-801-366-108</td> <td>Sequence 108, App</td>	1367	22	US-09-801-366-108	Sequence 108, App
39	199.5	8.9 <td>842</td> <td>20</td> <td>US-09-614-150-26857</td> <td>Sequence 55, Appl</td>	842	20	US-09-614-150-26857	Sequence 55, Appl
40	199.5	8.9 <td>842</td> <td>27</td> <td>US-60-191-637-26857</td> <td>Sequence 26857, A</td>	842	27	US-60-191-637-26857	Sequence 26857, A
41	199.5	8.9 <td>842</td> <td>27</td> <td>US-60-191-681-21466</td> <td>Sequence 21466, A</td>	842	27	US-60-191-681-21466	Sequence 21466, A
42	199.5	8.9 <td>864</td> <td>27</td> <td>US-09-614-150-40749</td> <td>Sequence 40749, A</td>	864	27	US-09-614-150-40749	Sequence 40749, A
43	199.5	8.9 <td>864</td> <td>27</td> <td>US-60-191-637-40374</td> <td>Sequence 40374, A</td>	864	27	US-60-191-637-40374	Sequence 40374, A
44	199.5	8.9 <td>864</td> <td>27</td> <td>US-60-191-700-351</td> <td>Sequence 351, App</td>	864	27	US-60-191-700-351	Sequence 351, App
45	193.5	8.7 <td>692</td> <td>20</td> <td>US-09-641-377-697</td> <td>Sequence 697, App</td>	692	20	US-09-641-377-697	Sequence 697, App

ALIGNMENTS

RESULT 1
PCT-US01-08631-39827
Sequence 39827, Application PC/TUS0108631
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-049
CURRENT APPLICATION NUMBER: PCT/US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO 39827
LENGTH: 801
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-08631-39827

Query Match 100.0%; Score 2249; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 11e-174; Indels 0; Gaps 0;
Matches 452; Conservative 0; Mismatches 0;

QY 1 GTLVTKVAPVSAAPKVSNGPRLPAPQIVAAKAPNTTITOPPANLQLPPTGLIKSNSGPL 60
DB 1 GTLVTKVAPVSAAPKVSNGPRLPAPQIVAAKAPNTTITOPPANLQLPPTGLIKSNSGPL 60

This Page Blank (uspto)

SEQUENCE CHARACTERISTICS:

LENGTH: 1140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match 6.9%; Score 156; DB 12; Length 1140;
Best Local Similarity 20.8%; Pred. No. 0.0054;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

QY 8 APVSAAPPVSSGPRLPAPQIVAVKAPNT-----TTIQFANLQLPFGVLI 53
DB 681 APYPPKEPAPYTPKPEAPYTPKETAAPYTPKGTAPYTPKLEPAAPYTPKKAPKELAPYT--- 737
QY 54 KNSGPMILVSPQOTVTRAEFTSNITSRAVPANPQVTKICVNPSSQLKKVAVTPVK 113
DB 738 --TKEPTSTSDKRAPYTPKGTAPYTPKPEAPYTPKETAAPYTPKGTATYTLKEPAAPYTPK 795
QY 114 KLAOIGTVVTVTPKPSVQSAVPTS---VVTVPCKPLNTVTVTLKPSLGASTPSNE 170
DB 796 KPAPKELAPYTPKGTSTSDKRAPYTPKETAAPYTPKPEAPYTPKPEAPYTPKPEAPYTPK 850
QY 171 PNLKENSAAVOINLSPTMLENVKCKNFMFLMLKLACSGSOSPMSGONVKKLYEQDLDA 230
DB 851 PPTTSEVSTPTTK-EPYTH-----KSP----- 873
QY 231 KIEAEFTRLKYLEKSSPOPLVFLKKSVALQLPLNSQSFIOCVQQTSSDMVIAT 290
DB 874 ---DEST---PELSAEPYTP-----KALENSPKPEGPPT 900
QY 291 CTYVTVTSPPVTVTVSSQSEKSIIVSGATAPRVSVOTLNPPLAGPVGAKAGVTVLHSGV 350
DB 901 TKTPAATKPEMTTKKDKTTERDL---RTTPETTT-----AAPKMTKETATTTTEKTT 949
QY 351 PTAATGTTAGTGL-----LQTSKPLVTSVANTVTVVSL--QPEK 388
DB 950 ESKITATTTQVSTTTQDTTPFKITTLTKTTLAPKVVTTKKITTTTEIMNKPEE 1003

RESULT 15
US-10-124-557-44

Sequence 44: Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-10-124-557-44

Query Match 6.9%; Score 156; DB 12; Length 1270;
Best Local Similarity 20.8%; Pred. No. 0.0062;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

QY 8 APVSAAPPVSSGPRLPAPQIVAVKAPNT-----TTIQFANLQLPFGVLI 53
DB 547 APYPPKEPAPYTPKPEAPYTPKETAAPYTPKGTAPYTPKLEPAAPYTPKKAPKELAPYT--- 603
QY 54 KNSGPMILVSPQOTVTRAEFTSNITSRAVPANPQVTKICVNPSSQLKKVAVTPVK 113
DB 604 --TKEPTSTSDKRAPYTPKGTAPYTPKPEAPYTPKETAAPYTPKGTATYTLKEPAAPYTPK 661
QY 114 KLAOIGTVVTVTPKPSVQSAVPTS---VVTVPCKPLNTVTVTLKPSLGASTPSNE 170
DB 662 KPAPKELAPYTPKGTSTSDKRAPYTPKETAAPYTPKPEAPYTPKPEAPYTPKPEAPYTPK 716
QY 171 PNLKENSAAVOINLSPTMLENVKCKNFMFLMLKLACSGSOSPMSGONVKKLYEQDLDA 230
DB 717 PPTTSEVSTPTTK-EPYTH-----KSP----- 739
QY 231 KIEAEFTRLKYLEKSSPOPLVFLKKSVALQLPLNSQSFIOCVQQTSSDMVIAT 290
DB 740 ---DEST---PELSAEPYTP-----KALENSPKPEGPPT 766
QY 291 CTYVTVTSPPVTVTVSSQSEKSIIVSGATAPRVSVOTLNPPLAGPVGAKAGVTVLHSGV 350
DB 767 TKTPAATKPEMTTKKDKTTERDL---RTTPETTT-----AAPKMTKETATTTTEKTT 815
QY 351 PTAATGTTAGTGL-----LQTSKPLVTSVANTVTVVSL--QPEK 388
DB 816 ESKITATTTQVSTTTQDTTPFKITTLTKTTLAPKVVTTKKITTTTEIMNKPEE 869

Search completed: February 16, 2003, 22:02:21
Job time : 23.9045 secs

OY 8 AVSAPKPVSSGRLPAPQIVAKAPNT-----TTIOFANLQLPSTVLI 53
DB 590 APTTPREPAPTPKEBAPTPKETAATPKGTAPTLKEBAPTPPKRPAKELAPTT---- 646
OY 54 KNSGGLMIVSQOYVTRAEITSNITSRAVPANPOTVACIVPNSSOLIKKAVANTPVK 113
DB 647 --TKRETSSTSDKPAETPKGTAPTPKEBAPTPPKETAATPKGTAPTLKEBAPTPPK 704
OY 114 KLAQIGTVVTVVPKPSVQSVAVPNTS---VTVPGKPLNTVTLKPSLSGASSTPSNE 170
DB 705 KAPKELAPTTTKGPTSTSDKPAETPKETAATPKETAATPKETAATPKETAATPKETA 759
OY 171 PUKAENSAVAOINLSPMLNENKCKNPLMLIKLACSGSPENGQVKKLVEQLDA 230
DB 760 PPTSEVSTPTTK--EPTTIH-----KSP----- 782
OY 231 KIEAEFTKRLVELKSSQPHLVPPLKSVVALROLNLSQSFIOQCVQOOTSMDVAT 290
DB 783 ---DEST---PELSAETPP-----KALENSKEGCVPT 809
OY 291 CTTVTVTSVTVTVSSQSEKSIIVSGATAPTVSVOTLNLPLAGVAKAGVTVLHSG 350
DB 810 TTPATKREMTTAKDKTTERDL---RTTPTTT-----AAPKMKETAATTEKTT 858
OY 351 PTAATGCTTAGTGL-----LQTSKPLVTSVANTVTVSL--QPERK 388
DB 859 ESKITATTTQVSTVQDTTPFKITTLKTTTLAPKVTITTKTTTTEIMNKPEE 912

RESULT 13

US-09-801-368-110
Sequence 110, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Moline, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Koyet, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 110
LENGTH: 1075
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-110

Query Match 6.9%; Score 156; DB 10; Length 1075;
Best Local Similarity 22.4%; Pred No. 0 005;

Matches 101; Conservative 63; Mismatches 170; Indels 116; Gaps 21;

OY 2 TLTVAVAPYAPKPVSSGRLPAPQIVAKAPN-----TTIOFANLQLPSTVLI 56
DB 427 TEVTITGNGQPTDET-----VIVIRPTSEGLITTTTE-----PWTGFTST 470
OY 57 SGPMLIV---SQOQ-----VTRAEITSNITSRAVP-----ANPQIVKICIVPNSSOLI 104

DB 471 STEMTVTCNGQPTDETVIIVIRPTSEGLISTTEPMTGTSTSTEVTLTGNGQPT 530
OY 105 KKVAV---TPVKKLAQIGTVVTVVPKPSVQSVAVANPSTVTVTPKPLN--TTLTKPS 160
DB 531 DEVIIVIRPTSE---GLITTTPEWGTGTSTSTEVTLTGNGQPTDETVIIVIRPT 586
OY 161 LGASSTPSNEPNKAKNSAVALI-----NLSPMLNENKCKNPLMLIK--LACSGS 211
DB 587 SEGILIRTEPMTGTSTSTEVTLTGNGQPT-----DEVIIVIRPTTALISS 637
OY 212 QSPFMQNVKKLVEQLDAKIEAEFTKRLVELKSSQPHLVP-----LKRSV- 262
DB 638 LSSSSQOITSSI-----TSSRPILTFYPSNCTSVISSVIS 674
OY 263 -ALROLPNOSPFIOQV-----QOTSDMVAATCTVTVTSPTVTVTVSSQ 309
DB 675 SSVTSSLVTSSTSSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSS 734
OY 310 SEKSIIVSGATAPTVSVOTLNLPLAGVAKAGVTVLHSGVPAATGTTAGTGLQTSK 369
DB 735 SSSSISSESPKSP--TSSSSLPV---TSATIGQETASSL--PRATTTTSE-----QTL 784
OY 370 PLVTSVANTVTVVSLQPERPVSGTAVTUS 399
DB 785 VTVTSCESHVCTESI--SSAIVSTATVTVS 812

RESULT 14

US-10-124-557-104
Sequence 104, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseer, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 104:

Db 832 ESKTATTTQVSTTTODTTPFKITTLKTTTLAKVTTTKTITTEIMKPEE 885

RESULT 11
US-10-124-557-74
Sequence 74, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match 6.9%; Score 156; DB 12; Length 1038;
Best Local Similarity 20.8%; Pred. No. 0.0048;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

QY 8 APVSAPPKVSSGRRLAPQVAVAKAPNT-----TTIQPPANLQDPGVL 53
Db 547 APTTPKEPAPPTTPKEPAPPTPKGTAPTTLKEPAPTPPKKPAKELAPTT--- 603
QY 54 KSNAGPLMLVSPQOTVRAETTSNITSRPAVPANPOTVAKICTVNSSSLIKKVAVPVK 113
Db 604 --TKPEPSTISDKPAPPTPKGTATTTKEPAPTPPKEPAPPTPKGTAPTTLKEPAPTPPK 661
QY 114 KLAQIGTTVTTVKPKSSVQSVAVPTS---VVTTPGKPLNTVTTLLKPSIGASSTPSNE 170
Db 662 KPAPKELAPTTTGGPSTSTSDKPAPPTPKKELAPTTPKEPAPT-TPKKP----APTTPETP 716

QY 171 PNKAENSAAVQINISPTMLENVKKCKNFMLMLKLACSSQSPKMGONKAKLYEQULDA 230
Db 717 PPTSEVSTPTTTR-EPTTH-----KSP----- 739

QY 231 KIEAEFTRKLYVELKSSPOPHLYFLKRSVALROLPLNSQSFIOQVOQTSSDMVAT 290
Db 740 ----DEST-----PELSAPTP-----KALENSPKPEPGVPT 766

QY 291 CTTVTTSPPVTTTVSSQSEKSIIVSGATAPRVSVQTLNPLAGPVGAKAGVYTLHSGV 350
Db 767 TKTPAATKPEMTTAKDKTERDL---RTTPETTT-----AAPKMTKEATTTTEKT 815

QY 351 PTAATGCTTACTGL-----LQTSKPLYSVANVTYVSL--QPEK 388
Db 816 ESKTATTTQVSTTTODTTPFKITTLKTTTLAKVTTTKTITTEIMKPEE 869

RESULT 12
US-10-124-557-58
Sequence 58, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match 6.9%; Score 156; DB 12; Length 1049;
Best Local Similarity 20.8%; Pred. No. 0.0048;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/124,557
  FILING DATE: 16-Apr-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/643,502
  FILING DATE: 18-JAN-1991
  APPLICATION NUMBER: US 07/546,114
  FILING DATE: 29-JUN-1990
  APPLICATION NUMBER: US 07/457,196
  FILING DATE: 29-DEC-1989
  APPLICATION NUMBER: US 07/390,901
  FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
  NAME: Casert, Luann
  REGISTRATION NUMBER: 31,822
  REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (617)876-1170
  TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 14:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 941 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-124-557-14

Query Match      6.9%; Score 156; DB 12; Length 941;
Best Local Similarity 20.8%; Pred. No. 0.0042;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

QY 8 AVSAPPKVSSGRRLPAPQIVAVKAPNT-----TTIOFPANLQLPFGTVLI 53
DB 482 APTTKREPAPPTPKREPAPPTPKETAPPTPKGTAPPTLKEPAPPTPKKPAKELAPTT--- 538
QY 54 KNSGSLMLVSPQOYVTRAEFTSNITSRPAVPANPQVAKICVPSNSSLIKKVAVTPEK 113
DB 539 --TKPEFTSTSDKPAPTPKGAPPTPKREPAPPTPKGAPPTPKGAPPTPKKPAKELAPTTK 596
QY 114 KLAQIGTVVTVTPKPSVQSAVAPTS---VVTVPGRPLNTVTTLKPSLIGASSPSNE 170
DB 597 KPAPKELAPPTTKGPTSTSDKPAPTPKETAPPTPKREPAPT-TPKKP---APTTPEPTP 651
QY 171 PNLKENSAAVQINSPTMLENVKCKNFMLMLIKLACSGSSPEKGNVKKLVEQLLDA 230
DB 652 PPTTSEVSTPTTK-EPTTIH-----KSP----- 674
QY 231 KIEAEFTKRLYLELKSSQPHLVPLFKSVVALRQLLPNSQSFIOOCVQOOTSMDVIAT 290
DB 675 ----DEST----PELSAETP-----KALENSPKKEGVP 701
QY 291 CTTVTTSPTVTVTVSSQSEKSIIVSGATAPRTVSYOTLNLAPGVAKAGVTVLHSGV 350
DB 702 TKTPATPKREMTTAKDKTTERDL---RTTPETTT-----AAPKMKETATTEKKT 750
QY 351 PTAANGTGTAGTGL-----LQTSKPLVTSVANTVTIVSL--QPEK 388
DB 751 ESKTATTQVSTTTQDTTPPKITTLKTTLLAPVTVTKKITTEIMNKEE 804
```

RESULT 10
US-10-124-557-84

Sequence 84, Application US/10124557
Patent No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

```
Gesner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
  ADDRESS: Genetics Institute, Inc.
  STREET: 87 Cambridgepark Drive
  CITY: Cambridge
  STATE: Massachusetts
  COUNTRY: U.S.A.
  ZIP: 02140
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/124,557
  FILING DATE: 16-Apr-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/643,502
  FILING DATE: 18-JAN-1991
  APPLICATION NUMBER: US 07/546,114
  FILING DATE: 29-JUN-1990
  APPLICATION NUMBER: US 07/457,196
  FILING DATE: 29-DEC-1989
  APPLICATION NUMBER: US 07/390,901
  FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
  NAME: Casert, Luann
  REGISTRATION NUMBER: 31,822
  REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (617)876-1170
  TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1022 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Query Match      6.9%; Score 156; DB 12; Length 1022;
Best Local Similarity 20.8%; Pred. No. 0.0047;
Matches 86; Conservative 40; Mismatches 164; Indels 124; Gaps 14;

QY 8 AVSAPPKVSSGRRLPAPQIVAVKAPNT-----TTIOFPANLQLPFGTVLI 53
DB 563 APTTKREPAPPTPKREPAPPTPKETAPPTPKGTAPPTLKEPAPPTPKKPAKELAPTT--- 619
QY 54 KNSGSLMLVSPQOYVTRAEFTSNITSRPAVPANPQVAKICVPSNSSLIKKVAVTPEK 113
DB 620 --TKPEFTSTSDKPAPTPKGAPPTPKREPAPPTPKGAPPTPKGAPPTPKKPAKELAPTTK 677
QY 114 KLAQIGTVVTVTPKPSVQSAVAPTS---VVTVPGRPLNTVTTLKPSLIGASSPSNE 170
DB 678 KPAPKELAPPTTKGPTSTSDKPAPTPKETAPPTPKREPAPT-TPKKP---APTTPEPTP 732
QY 171 PNLKENSAAVQINSPTMLENVKCKNFMLMLIKLACSGSSPEKGNVKKLVEQLLDA 230
DB 733 PPTTSEVSTPTTK-EPTTIH-----KSP----- 755
QY 231 KIEAEFTKRLYLELKSSQPHLVPLFKSVVALRQLLPNSQSFIOOCVQOOTSMDVIAT 290
DB 756 ----DEST----PELSAETP-----KALENSPKKEGVP 782
QY 291 CTTVTTSPTVTVTVSSQSEKSIIVSGATAPRTVSYOTLNLAPGVAKAGVTVLHSGV 350
DB 783 TKTPATPKREMTTAKDKTTERDL---RTTPETTT-----AAPKMKETATTEKKT 831
QY 351 PTAANGTGTAGTGL-----LQTSKPLVTSVANTVTIVSL--QPEK 388
```

FEATURE:
OTHER INFORMATION: MAP TO A1034555.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
OTHER INFORMATION: EST HUMAN HIT: A1117052.1, EVALUATE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUATE 3.00e-10
US-09-864-761-34248

Query Match 7.1%; Score 160; DB 10; Length 2665;
Best Local Similarity 24.1%; Pred. No. 0.0088;
Matches 87; Conservative 51; Mismatches 157; Indels 66; Gaps 14;

QY 9 PVSAPRVSAGPRUPAPQIVAVKAPNTTTIQFPANLQLPQIVLKSNGPLMVS--PQ 66
DB 1841 PVPVP-----SDPSIPPIPTLPSVTA-----AKLSPVASGCIHQSPPTKVTWITR 1887
QY 67 QVTTRAETTSNITSRAVPANPQTVKICTVPSNSOLIKKVAATP--YKLLAQIGTTVVT 124
DB 1888 QEPRRAOSTPS-----PALPPTKASDVDTSSSTL--KRLMDPKYVSAITSVTSVT 1939
QY 125 TVPKPSVSVAVPTSVTVTPGKPLNTVTLKPSLSGASSTPSNEPLKAENSAVAQIN 184
DB 1940 AIAEPYSAAPCLHEAPPPYDSKKPLEKTA--PVTNNSEIQASEVLVAADKEKAVP- 1996
QY 185 LSTPTMLNKKCKNFLAMLIKACSGSQSPKCONKVKLEQIDAKIEAEFTKRLYVE 244
DB 1997 IAKITSVISR-----MPVSIDLENSQKTKLAKPAQITLGVSA----- 2036
QY 245 LKSPQHLVPEFLKKSVMRLQLLPNSQSFIOQCVQOTSSDMVATCTTTVTSP--VVT 302
DB 2037 LUGLVNVLSP-----VNAKGPVKGSVTLKSLVSTPAPPVNVLKGPVNVLTGPVAVLT 2091
QY 303 TVVSSQSEKSIIVGATAPRTVSVOITLNPAGPVGAKAGVTLHSGVPTAATGTTAGT 362
DB 2092 TVNNAT-----VGTVANAPGTV-----NMAAASVNNATAAVTVTACAVTAAGSVATAT 2140
QY 363 G 363
DB 2141 G 2141

RESULT 8
US-09-801-368-106
Sequence 106, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amlir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19

PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 106
LENGTH: 1169
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106

Query Match 7.1%; Score 159.5; DB 10; Length 1169;
Best Local Similarity 21.0%; Pred. No. 0.0032;
Matches 105; Conservative 82; Mismatches 185; Indels 129; Gaps 22;

QY 4 VTKVAPVAPKPVSSGRLPAPQIVAVKAPNTTTIQF-PANLQLPQIVLKSNGPLM 62
DB 689 TVSSSVSTPRTSSSESSASVTLT-----PSTISEKPSPTMK-----TKVSISSPTNL 740
QY 63 VSPQQTTRAETTSNITSRAVP-----PANPQTVKICTVPSNSOLIKKVAATPV 112
DB 741 ITSYDTTSKOSTVGSSTVSLSISLIPSSYSASBDQIFHSSTIVSNGCALTSFSSSTKV 800
QY 113 K-----LAQIGTTVTVTPK-----PSSVQSVAVPNSVTVTPGKP 149
DB 801 SSSSESHRTPTTSSESGIKSGVELESTSTSFSEHETSTASTVQISQFVTPSSP 860
QY 150 LNTVTLKPSLSG---ASTPSNEPLKAENSAVAQINLSPTMLNKKCKNFLAMLIK 205
DB 861 ISTVA---PRSTGLNSQTESPTNSKKTMSSENSASV----- 893
QY 206 LACGSQSPKGVKVKLVLEQIDAKIEAE--ETR--KLVELKSSFPQHLVPELK 258
DB 894 MPSSSATSPKGT-----KVTSDTSSGFSRDRTVVMTSEPTSN-----EQ 936
QY 259 KSVVALQQLPNSQSFIOQCVQOTSSDMVATCTTTV-----TV-----SPVTTTVSS 307
DB 937 TTLTVSSCESNCS-----NTVSSANVSTATTINGITETVTCPLSATLELTVSK 989
QY 308 SQS-EKSIIVGATAPRTVSVOITLNP--LAGPVGAKAGVTLHSGVPTAATGTTAGTGL 364
DB 990 LSESEKTTTLTVTSCESGSETPASPAIVSPATVNDVTVTVSTWSPQATNKLAVSSD- 1048
QY 365 LQTSKPLVTSANVTTVSLQPEKPVYSGTAVTSLPAVTGEGTSGAICLPSKPVYSF 424
DB 1049 IENSKASFVSEAEKTSISRNNFVP--TSGTSTIEHT--TTSNASENSDMSASEAV 1106
QY 425 CWDHICKPVI-----GRP 437
DB 1107 SSKSVTNPVLISVQPRGTP 1127

RESULT 9
US-10-124-557-14
Sequence 14, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
APPLICANT: Clark, Stephen C.
APPLICANT: Jacobs, Kenneth
APPLICANT: Hewick, Rodney M.
APPLICANT: Geener, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

US-10-025-380-1068

Query Match	8.7%;	Score 195.5;	DB 9;	length 5179;
Best Local Similarity	-22.1%;	pred. No 7.7e-05;		
Matches 102;	Conservative 44;	Mismatches 208;	Indels 107;	Gaps 16;

```

0Y      2 TLVKKVAVPASPAPVSSGGRLPAPQIVAKAPNTTIOGPANLDLPQGVLLKNSGGLM 61
Db      1607 TTTTTTPTPTPTTSPSPPTPTTITP-PTSTITLLPRTTTPSSPRTTTTTTTPPPI- - - - -TTPSPPT 1661a
0Y      62 LVSPQQTVAET- - - - -TSTNISRPAVPANPQT- - - - -VKICTVPNSSQL 103
Db      1662 TTTTSPSPITTTTTPTPTPTTSPSSILTTTSSPPTTMTTSPRTTTPSSPITTTTTTPSSST- - 1719a
0Y      104 IKKAAVTPVKKKQIGTIVTVTPKSSVQSAVP- - -TSVTVYTPGKPLNTVYTLKPS 159
Db      1720 - - - - -TSPSPPTTMTTTPSPRTTTPSPPTTMTTTLPRTTTSSPLVTPPLDPSITTPSPSP 1773a
0Y      160 SLAGSSTP- - - - -SNEPNLKAENSAVQIN-LSPTMLENVKKCNFLAMDIK 205
Db      1774 STTTPPTPCVPLCMWGTGMLDSGKPNHKKGGDTLGDVCGRGMANI- - - - - 1821a
0Y      206 LACGSGOSP- - - - -EMQONVKALVEOLDLAKLEAEFTTKLYELKSSPQHLVPLKKS 260
Db      1822 -SCATATKMPVPVPIGLOLQOTVCDVSGVLCKKEDQ- - - - -KPG 1858a
0Y      261 VVALRQLLPNSOSPIQOCVOQTSNDWIATCT- - -TWTTSPPVTVTVSSSQSEKSIIVS 317
Db      1659 GVITPMACCLAMEINVOCCCECVOPTIMKTTTTTNTNRPRTTTPRTTITTTTTYVTPRTPTGQT 1918a
0Y      318 GATAPRTVSVQTLNPLAGPVGAK- - -AGVYTLHSVGPPAATAGT- - - - -TAGTGLL 365
Db      1919 PTTTTPITTTTT-TVVPRTPTPTGQTPTTTPTITTTTTYVTPRTPTPTGQDPTTPTPTITTTTTVT 1977a
0Y      366 QTSKPLVTSVANT- - -VTTVSLQPEKPVSSATNTLSLPAVT 404
Db      1978 PTPPTPTGQTPTPTTITTTTTVTPTPTPTGQTPTTPTPTT 2018a

```

```

RESULT 4
US-09-922-217-1068
: Sequence 1068, Application US/099222217
: Patent No. US20020076414A1
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Lodes, Michael J.
: APPLICANT: Secretist, Heather
: APPLICANT: Benson, Darin R.
: APPLICANT: Meagher, Madeline Joy
: APPLICANT: Stolk, John A.
: APPLICANT: Wang, Tongtong
: APPLICANT: Jiang, Yutlun
: APPLICANT: Smith, Carole Lynn
: APPLICANT: King, Gordon E.
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
: TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
: FILE REFERENCE: 210121.471C13
: CURRENT APPLICATION NUMBER: US/09/922.217
: CURRENT FILING DATE: 2001-08-03
: NUMBER OF SEQ ID NOS: 1124
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 1068
: LENGTH: 5179
: TYPE: PR1
: ORGANISM: Homo sapiens
US-09-922-217-1068

```

Query Match	8.7%;	Score 195.5;	DB 10;	Length 5179;
Best Local Similarity	22.1%;	Pred. No. 7.7e-05;		
Matches 102;	Conservative 44;	Mismatches 208;	Indels 107;	Gaps 16

```

0Y 2 TLKAVAVSAPRKVSSGPLPAPQIVAAKAPNTTIOGPANQLPBGVILKSSGGLM 61
Db 1607 TTTTTPTPTTSPPTTPTTTP--PTSTTTLPLPTTTPSPPTTTTTTTTTPTT--TTPSPPT 1661
0Y 62 LVSPQOTVTRAET-----TSTNTSHPAVPANPOT-----VKICVPMSSSOL 103
Db 1662 TTPSPSPITTTTTPTPTTTPSPPTTTPSPPTTMTPTSPPTTTPSPSITTTTTTPSST-- 1719
0Y 104 IKKAAVTPVKKLAOIGTTVTTTPKSSVSNAV-----TSVYVTPCKPLMTVTTLKPS 159
Db 1720 -----TPSPPTTTPPTTSPPTTTPSPPTTMTTLPLPTTSSPLTTPPTLPSTTPPTSPF 1773
0Y 160 SLGASSPT-----SNEPNLKAENSAVOJN--LSPMLTENYKCKNFPLAMLIK 205
Db 1774 STTTPPTPCVPLCMTWTGMLDSGRPNPKPGCDIELIGDVCGRMANI----- 1821
0Y 206 LACGSOSP-----EMGONAKLYEOLLDAKIEABEFTKRLVYELKSSPOHLPPLKKS 260
Db 1822 -SCHATMPDPVPIQOLQOTVCVSVGLICKNDQ-----KPG 1858
0Y 261 VVALROLLPNOSPTIOOCVOGTFSDMWIACF---TPTVTSPTVTTVSSSOSSEKSIIVS 317
Db 1859 GVIPMACTLNEIVOCCECVOTPTTMTTNTNPNPPTTTPITTTTTVTPPTPTPGTQ 1918
0Y 318 GATAPRTVSVOTLNPPLAGPVGAR---AGVYTLHSVGPRTATGT-----TAGTGL 365
Db 1919 PTTTPTTTTT--TVNPTPTPTGTOGPTTTPTTTTTTTTTVTPPTPTGTOPTPTPTTTTTVT 1977
0Y 366 QTSKPIVTSVANT--VTTTSLQPEKPRVSVSTATVLSLPAVT 404
Db 1978 PPTPTPGTQPTTPTTTTTTTTTVTPPTPTGTOPTPTTPTTT 2018

```

```

: RESULT 5
: US-09-833-263-1068
: Sequence 1068, Application US/09833263
: Patent No. US20020110547A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Ajun
: APPLICANT: Clapper, Jonathan D.
: APPLICANT: Stolk, John A.
: APPLICANT: Meagher, Madeline J.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
: FILE REFERENCE: 210121.471c12
: CURRENT APPLICATION NUMBER: US/09/833,263
: CURRENT FILING DATE: 2001-04-10
: NUMBER OF SEQ ID NOS: 1093
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1068
: LENGTH: 5179
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-833-263-1068

```

Query Match Score 8.7% ; Score 195.5; DB 10; Length 5179;
Best Local Similarity 22.1%; Pred. No. 7,7e-05;
Matches 102; Conservative 44; Mismatches 206; Indels 107; Gaps 16;

OY 2 TIVTVAVAPASAPKPKVSSGPRLPALPOLVAVKADNTTIGEPANLQCPCTVLIKNSGPIM 61
Db 1607 TTTTTPPPPTTTSSPPTTPIRP-PSTTTTLPTTSPSPPTTTTTPPP--TTSPSPPT 1661

OY 62 IVSPOOTVYRAET-----TSNITSRPVAVNPQT-----VKICTVPNSSQL 103
Db 1662 TTTSPPTTTTTTPTTPTTSSPITTTTSPPTTTMTTTPSPPTTTPSSPITTTTTPSST-- 1719

OY 104 IKKAVTPEPKLAOGTIVTVTPRPSVSQSAVP----TSVVYVTPGKLMTVTTLAKS 159
Db 1720 ----TTSPPPTMTTBSPTTTPSPPTTTMTTLTPPTTSSLTPTTTPPLPSPSTTTPPFSEF 1773

OY 160 SLGAASP-----SNPNILKAENSAAYOIN--LSPTMLENVKKCKNFAMLIK 205

US-09-893-519A-14

Query Match 22.8%; Score 513; DB 9; Length 1023;
Best Local Similarity 33.4%; Pred. No. 1.3e-27;
Matches 154; Conservative 62; Mismatches 127; Indels 118; Gaps 19;

QY 13 PRVSSG-----PRLPAQIVAKAPNTTIGPPANIQLPBGTVLKSNSGPI 61
DB 414 PRATSGIGATLPLVLABRLPQ-----PQNPNIQ---NFQLPBGVLVSENGQL 464
QY 62 LVSPOQTVR-----AETSNTSRPAPNPQVTKICTVPRSSQLIKKAVATVYKL 115
DB 465 MI-PQALAKOMQAQAQAQOTTMAPRPTSPAPVOISTVAPGPTIAR-QVTP-----518
QY 116 AIGITVTVTPKPSVQSAVPTSVYTPGKPLNTYT--TLKPSLSASS-----TPS 168
DB 519 ----TIIIOV---SQAQITVQPSATLQSPGQPOLVIGGAQATSLGATATVQGTQ 571
QY 169 NE-PNLKANSAAVOINISPTMLENVKCKNPLAMLIKACSGSGSPKMGQNKVLEOL 227
DB 572 RVPVGGATTSSAATE-----TMEVYKCKNFKLSTLIKIASGKQSTETAAVKELOV 625
QY 228 LKAKTEAEFTKRLVLEKSSPOPHVPLKKSVALRQLRPSQSFIOQCVO-----TS 283
DB 626 LQKTEAEFTKRLVLEKSSPOPHVPLKKSVALRQLRPSQSAFIOQSOQPPPPS 685
QY 284 SDMVATCTTWTSPVTTTSSQSEKSIYSAGTAPRVSQTLNPLAGPVAKAGV 343
DB 686 Q-----ATTALVAVLSSSVQRTAGKTAATVTSALQPPVLSL-----722
QY 344 VTLHSVGPPTAATGTTAGTGLQTSKPLVTSANVTVTSLOPEKPVSGTAVTSLP 403
DB 723 -----TOPVGVGKQSGQPPVLIQ-----GPPKP-----GALIRPQV 756
QY 404 TGETSGAALCLPSVKPVVSFCMDHICKPVIGTPVOIKLAQ 444
DB 757 TLTOT-----PMVALRQPH-NRIMLTTPQOVNLS 785

RESULT 2

US-09-801-368-108
Sequence 108 Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Milne, Todd
APPLICANT: Maxon, Mary
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 108
LENGTH: 1367
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Query Match 8.9%; Score 200.5; DB 10; Length 1367;
Best Local Similarity 22.2%; Pred. No. 6e-06;
Matches 113; Conservative 69; Mismatches 225; Indels 101; Gaps 17;

QY 2 TLVTKVAPVAPKVSQSGRLPAP-----QIVANKAPNTTIGPPANIQLPBGTVL 54
DB 618 TRESSAPVYVSTRESSAPVPPSSSTRESSAPVPPSSSTRESSAPVPPSSSTRE 677
QY 55 SNGPMLVSPQOT---VRAETSNITSRPANPQVTKICTVNNSQLIKKAV-- 109
DB 678 SSSAPVSTRESSAPVSTRESSAPVPPSSSTRESSAPVPPSSSTRESSAPV 737
QY 110 -PPVKIAQIGITVVT-----VPPRSSVQS-----VAVPTSVYVTPGKPLNT- 152
DB 738 PTPSSSTRESSAPVSTRESSAPVPPSSSTRESSAPVPPSSSTRESSAPVPP 797
QY 153 -----YTLKPSLSGASTSPNEPNLKAENSAVOILSTMLENVKCKNF 201
DB 798 SSSSTRESSAPVPPSSSTRESSAPVPPSSSTRESSAPVPPSSSTRESSAPV 851
QY 202 MLKILACSGSOS-----PEMGQNKVLEQLDAKTEAEFTKRLVLEKSSPQPH 254
DB 852 --APVSSSTRESSAPVPPSSSTRESSAPVPPSSSTRESSAPVPPSSSTRE 904
QY 255 PFLKKSVALRQLRPSQSFIOQCVOQTSADVIACTTTVTSP---VTTTVSSQSR 311
DB 905 -----YPSQRT-ETVSSTETETIVPKRTTSTVTPPTTTTITTVCGSTGN 949
QY 312 KSITVSGATAPRV---SVQTLNPLAGPVAKAGVTLHSVGPPTAATGTTAGG--LL 367
DB 950 SAGETTSVSCSPKVTVTTPPTTTTSTSTTTTITTVCGSTGNSAGETTSV 1009
QY 368 SKPLTVSVANT-----VTVTSLOPEKPVSGTAVTSLPANT--FGTSGAIC-----414
DB 1010 TVPCSTSPSTSESTTSTPTTPTVTVSTTVTTSTSTKAGGELITTVTKINIP 1069
QY 415 -----LPSVKPVVSFCMDHICKPVIGT 436
DB 1070 LTTIAPPTSVTVVNTFTTITTVCGST 1097

RESULT 3

US-10-025-380-1068
Sequence 1068 Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aljun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skelky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1068
LENGTH: 5179
TYPE: PRT
ORGANISM: Homo sapiens

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 8.9045 Seconds
(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452
Perfect score: 2249
Sequence: 1 GRLVTKVAPVAPKSSCP.....VICTPVQIKLQEPVLISOP 452

Scoring table: BLOSUM62

Gaped 10.0 , Gapped 0.5

Searched: 140259 seqs, 2554876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB pep:*
2: /cgn2_6/ptodata/2/pubppa/PC01_NEW_PUB pep:*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB pep:*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB pep:*
6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/2/pubppa/PC01_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB pep:*
10: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB pep:*
11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB pep:*
12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB pep:*
14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	513	22.8	1023	9 US-09-893-519A-14	Sequence 14, Appl
2	200.5	8.9	1367	10 US-09-801-368-108	Sequence 108, Appl
3	195.5	8.7	5179	9 US-10-025-380-1068	Sequence 1068, Ap
4	195.5	8.7	5179	10 US-09-922-217-1068	Sequence 1068, Ap
5	195.5	8.7	5179	10 US-09-833-263-1068	Sequence 114, App
6	172	7.6	1322	10 US-09-801-368-114	Sequence 34248, A
7	160	7.1	2665	10 US-09-864-761-34248	Sequence 106, App
8	159.5	7.1	1169	10 US-09-801-368-106	Sequence 14, Appl
9	156	6.9	941	12 US-10-124-557-14	Sequence 84, Appl
10	156	6.9	1022	12 US-10-124-557-84	Sequence 74, Appl
11	156	6.9	1038	12 US-10-124-557-74	Sequence 58, Appl
12	156	6.9	1049	12 US-10-124-557-58	Sequence 110, App
13	156	6.9	1075	10 US-09-801-368-110	Sequence 104, App
14	156	6.9	1140	12 US-10-124-557-104	Sequence 44, Appl
15	156	6.9	1270	12 US-10-124-557-44	Sequence 42, Appl
16	156	6.9	1311	12 US-10-124-557-42	Sequence 142, App
17	156	6.9	1314	12 US-10-124-557-142	Sequence 50, Appl
18	156	6.9	1314	12 US-10-124-557-50	Sequence 46, Appl
19	156	6.9	1320	12 US-10-124-557-46	

20	156	6.9	1320	12 US-10-124-557-60	Sequence 60, Appl
21	156	6.9	1354	12 US-10-124-557-48	Sequence 48, Appl
22	156	6.9	1361	12 US-10-124-557-40	Sequence 40, Appl
23	156	6.9	1363	12 US-10-124-557-52	Sequence 52, Appl
24	156	6.9	1404	12 US-10-124-557-2	Sequence 2, Appl1
25	156	6.9	1404	12 US-10-124-557-62	Sequence 62, Appl
26	153	6.8	1601	10 US-09-862-027-40	Sequence 40, Appl
27	152.5	6.8	386	10 US-09-864-761-35720	Sequence 35720, A
28	150	6.7	1056	9 US-10-161-510-10	Sequence 10, Appl
29	150	6.7	1537	10 US-09-801-368-104	Sequence 104, Appl
30	145	6.4	688	10 US-09-864-761-36047	Sequence 36047, A
31	142.5	6.3	2586	10 US-09-905-129-11	Sequence 11, Appl
32	142.5	6.3	2586	10 US-09-905-129-14	Sequence 14, Appl
33	142.5	6.3	2586	10 US-09-905-129-11	Sequence 11, Appl
34	142.5	6.3	2586	10 US-09-905-129-11	Sequence 11, Appl
35	142.5	6.3	2587	10 US-09-905-129-16	Sequence 16, Appl
36	142.5	6.3	2587	10 US-09-905-129-16	Sequence 16, Appl
37	142.5	6.3	2589	10 US-09-905-129-24	Sequence 24, Appl
38	141	6.3	2597	10 US-09-905-129-2	Sequence 2, Appl1
39	141	6.3	2597	10 US-09-905-129-10	Sequence 10, Appl
40	141	6.3	2597	10 US-09-905-129-13	Sequence 13, Appl
41	141	6.3	2597	10 US-09-905-129-13	Sequence 13, Appl
42	141	6.3	2597	10 US-09-905-129-13	Sequence 13, Appl
43	141	6.3	2597	10 US-09-905-129-13	Sequence 13, Appl
44	139	6.2	2005	10 US-09-735-3678-3	Sequence 3, Appl1
45	139	6.2	2063	10 US-09-735-3678-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-893-519A-14
Sequence 14, Application US/09893519A
Publication No. US20030027243A1
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURDMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOKARINTSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tairig
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)...(1023)

This Page Blank (uspto)

Db 255 SLANO-OLPPACAROLSKRLTLQOF--GNDISPEIGERYRLVLGVNSTLTIE 311
 QY 236 EFTKRLVELKSSPOPHLYPFLKKSVAL-ROLLPNSQSFIO---QCVOQISSDMVITFC 291
 Db 312 EFKSLLEATNPPLRPFLVPLKANLPLORELLHCAFLAKONPAQYLAOHQOLLDAST 371
 QY 292 TTTVTTSFV 301
 Db 372 TSPVDSSELL 381

RESULT 15
 US-08-325-267A-4
 : Sequence 4, Application US/08325267A
 : Patent No. 5585271
 : GENERAL INFORMATION:
 : APPLICANT: WATARI, JUNJI
 : APPLICANT: TAKATA, YOSHIHIRO
 : APPLICANT: OGAWA, MASAHIRO
 : APPLICANT: ONNELA, MERJA
 : APPLICANT: KERANEN, SIKKA
 : TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
 : TITLE OF INVENTION: CONTAINING THEM
 : NUMBER OF SEQUENCES: 7
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
 : STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
 : CITY: ARLINGTON
 : STATE: VA
 : COUNTRY: USA
 : ZIP: 22202
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/325.267A
 : FILING DATE: 18-NOV-1994
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP PCT/JP94/00290
 : FILING DATE: 24-FEB-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 38871/1993
 : FILING DATE: 26-FEB-1993
 : ATTORNEY/AGENT INFORMATION:
 : NAME: OBLON, NORMAN F
 : REGISTRATION NUMBER: 24,618
 : REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 703-413-3000
 : TELEFAX: 703-413-2220
 : TELEX: 248855 OPAT UR
 : INFORMATION FOR SRO ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 862 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-325-267A-4

Query Match 7.3%; Score 163.5; DB 1; Length 862;
 Best Local Similarity 21.1%; Pred. No. 1.4e-05;
 Matches 102; Conservative 67; Mismatches 156; Indels 159; Gaps 21;

QY 11 SAPKXSSGRLLPAPQIVAAVAPNTTIQF---PANQLPPTGVILIKNSGPL----- 60
 Db 199 SLPEITGTYVMAGYGYPMKVVSNAVSWGTDLISVTLPGDTGVYSDDFEGYVVSFDDDL 258
 QY 61 -----MLVSPQOQVTVTRAET-----TSNITSRPAVPANPOTVTKICTVPS 99

Db 259 SOSNCTVPDPNSYAVSTTTTTEPMTGFTSTSTEMTIVTGNGVPTD-ETVIVIRTPPT 317
 QY 100 SSQLIKVAVTPVKRLAQIGTIVTTPKPSVQSAVPTSVTVTPGKPLN-TVTTIK- 157
 Db 318 AS-----FTITTEPMTGFTSTSTEMTIVTGNGVPTD-ETVIVIR 359
 QY 158 PSSLGASGPNSEF-----NLKAENSAVQINLSPTMLENVKKKNFLAMIKIACSGS 211
 Db 360 PTSEGLVTT-TTEPMTGFTSTSTEMTIVTGNGVPTD-ETVIVKT-PTNAISSLS 416
 QY 212 QSPENGQVKKLYEQLDLAKIEAEFTKLYELKSSPOPHLYPFLKKSVALROLLPNS 271
 Db 417 SSGQITTSI-----TSSRPITTPPPS-----NG 440
 QY 272 QSFIOQCVOQISSDMVIACTTV-TTSPVTTTVSS-----SQSEKSIIV- 316
 Db 441 TSVI-----SSSVISSVTSLSFTSPVSSVSSSTTSTSTSTSTSTSTSTSTSTST 493
 QY 317 --SGATAPRTVSQTLNPLAGPYGAKAGVYVTHASGPTNAAGCTTAGTGLQTSKPLVTS 374
 Db 494 STSGSSESTSS-----AGSVSSSSFTSSSSKSPYSS-----SLPLVTS 535
 QY 375 VANTVTVSLQPEKRPVSGTAVTLSLPVTGFTSGAALCLPSVKPVVSCMDHICKRPVI 434
 Db 536 ATTSQETAS-----SLPATTTTKTSEQTLV-----TVTSCSHVCTESI 575
 QY 435 GTPV 438
 Db 576 SPAL 579

Search completed: February 16, 2003, 22:02:50
 Job time : 20.6146 secs

QY 397 TSLPAP 403
DB 677 TTPLPDI 683

RESULT 13

US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480,76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-6

Query Match 7.6%; Score 171.5; DB 3; Length 1721;
Best Local Similarity 21.5%; Pred. No. 8.3e-06;
Matches 92; Conservative 50; Mismatches 198; Indels 87; Gaps 15;

QY 35 TTTTTPANLQPLPGTVLTKNSGMLVSPQOTYRAETTSNITSRPAVPANPQTVKIC 94
DB 286 TTTTTPANLQPLPGTVLTKNSGMLVSPQOTYRAETTSNITSRPAVPANPQTVKIC 94
QY 95 TYPNSSOLIKKVAATPVAKKLAQIGTVTVTPKPSVSVAVPVSIVTVTGKPLNTVT 154
DB 346 TTTTTPANLQPLPGTVLTKNSGMLVSPQOTYRAETTSNITSRPAVPANPQTVKIC 94
QY 155 TLKPSLSGASTPS-----NEPNLKAENSA-----VOINLSPMLDN-- 192
DB 399 TTAATTTTTSSESVIAKPDKMCWLEKNGCEAKAGATYVVGKGRIEGMAFTMLPND 458
QY 193 -----VKCKNFLAMLIKILACSGSGSPENGQNVKLVLEQLDAKIE--AEFTKLY 242
DB 459 THVRFRFKVDGNTISVCRKAGKLEFP-----DRSLDTIPVAGHNSCII 508
QY 243 VELKSSPOHVLPLKKSVALROLLPNSQF-----IQCV-----OOTSMMVIA 289

DB 509 VGVSGDGKIHVSYPGSKDVLISAPIQSELEFNEVYCDCTAKYGAHSGQYSDAEVTT 568
QY 290 TC---TTTWTSPVYTTTSSOSEKSIIVSGATADRTVSQTLNPLAGVGAAGVTL 346
DB 569 TTAATTTTTSSESVIAKPDKMCWLEKNGCEAKAGATYVVGKGRIEGMAFTMLPND 458
QY 347 HSYGPTMATGTTAGTGLLOTS---KPLVTSVAN--TTVVSLQPEKPVSG---TAV 396
DB 617 TTKKPTTTTTSSESVIAKPDKMCWLEKNGCEAKAGATYVVGKGRIEGMAFTMLPND 458
QY 397 TSLPAP 403
DB 677 TTPLPDI 683

RESULT 14

US-08-244-189-2
; Sequence 2, Application US/08244189
; Patent No. 5580727
; GENERAL INFORMATION:
; APPLICANT: Ohki, Misao
; APPLICANT: Kikuchi, Kimiko
; APPLICANT: Miyoshi, Hiroyuki
; TITLE OF INVENTION: No. 5580727el DNAs, Polypeptides Encoded Thereby
; TITLE OF INVENTION: and Methods for Detecting the DNAs and Polypeptides
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,189
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 760-183P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-244-189-2

Query Match 7.3%; Score 164.5; DB 1; Length 752;
Best Local Similarity 28.8%; Pred. No. 9.7e-06;
Matches 72; Conservative 41; Mismatches 108; Indels 29; Gaps 11;

QY 61 MLVSPQOTYRAETTSNITSRPAVPANPQTVKICYPNSSOLIKKVAATPVAKKLAQIGT 120
DB 152 VFTNPQVATYHRAIKITVDGPREPN-RTERKHSITPDPVDTQSRLLP----- 201
QY 121 TTTTTPANLQPLPGTVLTKNSGMLVSPQOTYRAETTSNITSRPAVPANPQTVKIC 94
DB 202 ---PMPPTTQG-APRISSTPT---TLTNGTSHSPALNAGAPPPGCFNSGSSSS 254
QY 177 NSANVOINISPTM-LENVKKCNFLAMLIKILACSGSGSPENGQNVKLVLEQLDAKIEAE 235

PCT-US93-11721-5

Query Match 9.1%; Score 204; DB 5; Length 2035;

Best Local Similarity 23.5%; Pred. No. 1.7e-08;

Matches 119; Conservative 66; Mismatches 188; Indels 134; Gaps 24;

QY 10 VSAPKPVSSGPRLPAPQIVAVKAPNTTTIOFPANLQLPFGYVLKINSGLMLVSPQTV 69
DB 537 ICSIPMSGMAALAAAAATOKIPSSA--PVLVSPAGATTVKT-----MAVPTGTTT 588
QY 70 TSEETSNITSRPAPVAPNTQKICVPNSSGLKAVVAPVTKKIAQIGTV----- 122
DB 589 LPA--TVKASSPVMVSNAT-----RMLKTA--AAVGTSSAKTNTS 628
QY 123 ---VTVPRSSVQSAVPTSVTVTPGKPLNTVTLK-PSL-GASTSPNEPNL----- 173
DB 629 TPRTITVHSGTV-TVAAQAAQVTVTVGGVTKITILVKPISVPGSALISNIGKMSV 687
QY 174 --KAENSAAVQINLSFTMLENVKCKNPL--AMLKILACSGSOSP-----EMCONVK 221
DB 688 QTRPVQTSVNTQASGTPYTOIOTKGPAGTILKLVTSADGKPTTITTTQASGATK 747
QY 222 KIVEOLDLAKIEEFTKRLVLELKSSPQHLVPLFKKSVVALRQLPMSQSFIOQVQ 281
DB 748 PTLTIGI-----SSVSPSTT--KPGTTTITKTPMSAITTOAGANG 785
QY 282 TSSDMVIATCTTVT-----SPVVT-----TVSSSQSEKSIIVSGA-----TA 321
DB 786 VNSSGIKRPITITTKVMTSGNAPAKITITAVPKIATHGQGVYQVVLKAPGPGPT 845
QY 322 PRT-----VSQTLNPLAGPVAK--AGVTLHSGVPTAATGTTACTGLLQTS 368
DB 846 LRTVPMGVRLVTPVVSIVKPAVTLVYKGTGVTTLCTGTVST--SLAGAGSHS 903
QY 369 KPLVTSVA--NTVTYSIQPEKPVYSTAVTLSLPVTGEGTSGAICLPSVYVVSFCM 426
DB 904 ASLAPITTLTGITATLSSQ---VINPTAVISAOTTLTAAGLTTPTTMMQPV----- 954
QY 427 DHICKPVIGTPVOIKL-AQPGVLSOP 452
DB 955 -----SQPTQVTLITAPSGVEAQP 973

RESULT 9
US-09-579-181-2

; Sequence 2, Application US/09579181

; Patent No. 6365372

; GENERAL INFORMATION:

; APPLICANT: Chivria, John

; APPLICANT: Yaciuk, Peter

; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)

; FILE REFERENCE: 16153-4247

; CURRENT APPLICATION NUMBER: US/09/579,181

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: 60/136,620

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 2

; LENGTH: 2972

; TYPE: PRT

; ORGANISM: Human

US-09-579-181-2

Query Match 7.8%; Score 176.5; DB 4; Length 2972;

Best Local Similarity 22.4%; Pred. No. 6.9e-06;

Matches 123; Conservative 56; Mismatches 196; Indels 175; Gaps 26;

QY 9 PVSAPPKVVSSGPRL-----PAPQIVAVKAPNTTTIOFPANLQ-----P 47
DB 826 PVRPPP-----GPELSAQPTPGVPQVL-----PASLMSVSPAGPPLIPASRP 869
QY 48 PCTVL---IKNSG--PLMLVSPQOTVTRAETTSNITSRPV-----PANQTVKICTV 96

DB 870 PGVLLPLQPNSSGLPQVLPSPGLVLSG-----TSRPPTTSLKPTPAPVRLSPA 922
QY 97 -PNSSQLIKKVAVTPVKKLAQIGTVTVTPRPSSVQSAVPTSVYTPGKPLNTVTT 155
DB 923 PPGSSSLKPLVLPVPGYTPPAAATTTSTTATATTAVPAT-----PAPORLI 973
QY 156 LKPS-----SLGASTSPNEPNLKAENSAVQINLSFTMLENVKCKNPLAMLI 204
DB 974 LSPDMQARLPSSGVEVSIQGLASLAQRPVANAGSKRLFTFOIGN----- 1017
QY 205 KLAGSGSQ-----SPENQONVKLVEOLDLAKIEEFTKRLVLELKSSPQHLVPL 257
DB 1018 KLTLTGAQVROLAVGQPRPLQMPPTVWNNMGVYKIVVRAQPRD-----GLTPVPPLAPAP 1072
QY 258 KKSVALRQLPLPSQSFIOQCVQOTSSDMVIANCTTVTTSPTVTT-----VSSQ 309
DB 1073 RPPSSGLPAVLN-----PRPTLPGRLPPTVLTGTAAPAPPTPLVRLKLVHSPS 1123
QY 310 SEKSIIVSGATAPRTVS-----VQTLNPLAGPVAKAGV-----V 344
DB 1124 PEVSASAPGA-APLTSSPLHVPSSLPGPASSMPPIPNSSPLASPVSVSVPLSSSLPI 1182
QY 345 TLHSGPPTAATGTT-----AGTGLL-QTSKPLVTSYANTVTVSLQPEKRVVS 392
DB 1183 SVPTTLPAAPASAPLTIPISAPLTVSASGPAALLTSVTPPLAPVPAAPGPPSLQPSGSPS 1242
QY 393 GTRAVTSL---PAVTGEGTSGAICL-----PSVKPVVSFCMDHICKPVIGTPVOIKL 442
DB 1243 ASALVTGLATPDSLSSQTPGHPDLLAPTSSHAVGINSVA---PACSPVL-VPAS-AL 1296
QY 443 AOPGPVLSOP 452
DB 1297 ASPFPASRP 1306

RESULT 10
US-09-579-181-1

; Sequence 1, Application US/09579181

; Patent No. 6365372

; GENERAL INFORMATION:

; APPLICANT: Chivria, John

; APPLICANT: Yaciuk, Peter

; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)

; FILE REFERENCE: 16153-4247

; CURRENT APPLICATION NUMBER: US/09/579,181

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: 60/136,620

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; LENGTH: 3118

; TYPE: PRT

; ORGANISM: Human

US-09-579-181-1

Query Match 7.8%; Score 176.5; DB 4; Length 3118;

Best Local Similarity 22.4%; Pred. No. 7.5e-06;

Matches 123; Conservative 56; Mismatches 196; Indels 175; Gaps 26;

QY 9 PVSAPPKVVSSGPRL-----PAPQIVAVKAPNTTTIOFPANLQ-----P 47
DB 972 PVRPPP-----GPELSAQPTPGVPQVL-----PASLMSVSPAGPPLIPASRP 1015
QY 48 PCTVL---IKNSG--PLMLVSPQOTVTRAETTSNITSRPV-----PANQTVKICTV 96
DB 1016 PGVLLPLQPNSSGLPQVLPSPGLVLSG-----TSRPPTTSLKPTPAPVRLSPA 1068
QY 97 -PNSSQLIKKVAVTPVKKLAQIGTVTVTPRPSSVQSAVPTSVYTPGKPLNTVTT 155
DB 1069 PPGSSSLKPLVLPVPGYTPPAAATTTSTTATATTAVPAT-----PAPORLI 1119

Db 748 PTLIGT-----SSVSPST---KPGTTITIKTIPMSAITTQAGATG 785
QY 282 TSDMVIATCTTIVT-----SPVVT-----TVSSSQSEKSIIVGCA-----TA 321
Db 786 VTSSPGIKSPITITITKMTSGTGAPAKIITAVPKIATGHQOGVTVYVLGABGQPGPTI 845
QY 322 PRT-----VSQTLNPLAGPVGAK--AGVTLHSVGPITAAAGTTAGTGLQTS 368
Db 846 LRTVPMGVRLLVPTVSAVKPAVTTLVKGTGTGTVTGTVST--SLAGAGHSTS 903
QY 369 KPLVTSVA--NTVTVTSIQPEKPVVSGTAVTSLPAVTFGETSGAALCLPSVKPVVSRGW 426
Db 904 ASLATPITTLTGITITLSQ-----VINPAAIVSAAGTTLTAAGGLTPTTMOV----- 954
QY 427 DHICKPVIPTVQIKL-AQPGPVLSPQ 452
Db 955 -----SOPTQVTLITAPSGVEAQP 973

RESULT 7
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,703
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-2/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-393-703-5

Query Match 9.1%; Score 204; DB 1; Length 2035;
Best Local Similarity 23.5%; Pred. No. 1.7e-08;
Matches 119; Conservative 66; Mismatches 188; Indels 134; Gaps 24;

QY 10 VSAPRKVSSGPRLPAPDIYAVKAPNTTITQIPANLQLPPTGVILIKSNSGRLMLVSPQGV 69
Db 537 IGSSPQMSGMAALAAAAAATOKIPSSA---PVLVSPAGTTIVKT-----MAVTPGTTT 588
QY 70 TRAETTNISRPVAVPAPQVTKICTVNPSSQLIKKVAIVPVAKKLAQIGTTV----- 122

Db 569 LPA--TVAVASSPVAVSNPAT-----RMKLTAA-----AQVGSVSSAINTS 628
QY 123 ---VTVPKPSSVOSVAAVTSVTVTPGKPLNTVTLK--PSSL--GASSTPSENPUL----- 173
Db 629 TRPITTVKSGTV--TVAAQAVVTVTVGCVTKITILVKSPLSVPGSALJNLKVMYV 687
QY 174 --KAENSAVQINISPTMLENVKCKCNFL--AMLIKLACSSQSP-----EKQNVK 221
Db 688 QTKPVQTSVAVTGOASTGCVPTQIIQTKGPLPAGTILKLVTSADGKRTTITTTTQASGACTQ 747
QY 222 KLVEQLDAKLEAEFFTKLVELKSSPQPHVPEPLKSVVALROLNLSQSFLOQCQVQ 281
Db 748 PTLIGT-----SSVSPST---KPGTTITIKTIPMSAITTQAGATG 785
QY 282 TSDMVIATCTTIVT-----SPVVT-----TVSSSQSEKSIIVGCA-----TA 321
Db 786 VTSSPGIKSPITITITKMTSGTGAPAKIITAVPKIATGHQOGVTVYVLGABGQPGPTI 845
QY 322 PRT-----VSQTLNPLAGPVGAK--AGVTLHSVGPITAAAGTTAGTGLQTS 368
Db 846 LRTVPMGVRLLVPTVSAVKPAVTTLVKGTGTGTVTGTVST--SLAGAGHSTS 903
QY 369 KPLVTSVA--NTVTVTSIQPEKPVVSGTAVTSLPAVTFGETSGAALCLPSVKPVVSRGW 426
Db 904 ASLATPITTLTGITITLSQ-----VINPAAIVSAAGTTLTAAGGLTPTTMOV----- 954
QY 427 DHICKPVIPTVQIKL-AQPGPVLSPQ 452
Db 955 -----SOPTQVTLITAPSGVEAQP 973

RESULT 8
PCT-US93-11721-5
; Sequence 5, Application PC/TUS9311721
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11721
; FILING DATE: 03-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-57503-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide


```

APPLICANT: Dynlact, Brian D.
APPLICANT: Hoeey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299-
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-2

Query Match 10.7%; Score 241.5; DB 1; Length 921;
Best Local Similarity 24.8%; Pred. No. 3.2e-12;
Matches 112; Conservative 65; Mismatches 163; Indels 111; Gaps 17;

QY 25 PQIVANKAINTTIOIPANILQIPGIVLILKNSGPIML-----VSFOQVYTRAETTS 76
DB 133 POSPTLSTLTNGOTPA-----LVKTDNGFQLRVGTTGPTVOTITNTSNN 184
QY 77 NITSRAVPANPQVAKICIVPNSSQ-----LIRKVAVTPVKLAQIGTVVTV 127
DB 185 NITSTTNHFTTQ-IRLOVPPAAASMTNTTATSNITVNSVASSGYANSSQPHLITLQ 243
QY 128 KSSVQSAVPTSVTVTGKPLNTVTLKPSLSGASSTPSEPNLKAENSAVAQINLSP 187
DB 244 APOLPQITQITIPAOSSQOQVNNVSSAGATATAVSSFTA-----ATT 287
QY 188 TULEANK-KCKNPLMLILACSGSPPMQONVKKLVQQLDAKLEAEFTKLYELK 246
DB 288 TOGNTKKECKRFFLANLIL--STREPKPEKRVRLTLOELVNAVPEPEFCDRLERLIN 345
QY 247 SSPQHLVPELKKSVVALROL-----LPNSOSFIO--- 276
DB 346 ASBPOLIGLFLKSLPLRLQALYTKELVLEGKPPROHVLGLAGLSQOLPKIOAQIRP 405
QY 277 ----QCVOQTSSDMVIACTTTVTTSPPVTVVSSQSEKSIIVSGATAPRTVS---VOT 329
DB 406 PSQTTTIGQTVRMI--TFNALGTPTPTIGHITTSIKOPN---IRLPYAPRLVNTGIGRT 460
QY 330 LNPAGPVAKAGVYVTLHSGVPTAATGCTTAGTGLQTSKPLVTSVANTVTVSIQPEK 389

```

```

DB 461 QIP-SLQVPGQANIVQIR--GPOHAQLOFTGVSQVIRATTRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTLSPAVTFGTSGAALICPSVKP 420
DB 507 -----KLTAVKVGQTOIKAI-TPSLHP 527

RESULT 6
US-08-046-585-5
Sequence 5, Application US/08046585
Patent No. 5453362
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
APPLICANT: Herr, Winship
TITLE OF INVENTION: A NOVEL EKARIOTIC TRANSCRIPTION PROTEIN:
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/046,585
FILING DATE: 12-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299-
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-046-585-5

Query Match 9.1%; Score 204; DB 1; Length 2035;
Best Local Similarity 23.5%; Pred. No. 1.7e-08;
Matches 119; Conservative 66; Mismatches 188; Indels 134; Gaps 24;

QY 10 VSAIPVSSGPRLPAPQIVAVKAPNTTIOIPANLQIPGIVLILKNSGPIMLVSPQTV 69
DB 537 IGSSPQSGMALAAALAAATOKIPSSA---PVLVSVPAGTIVTKT-----NAVTPGTTT 588
QY 70 TRAEFTSNITSRAVPANPQVAKICIVPNSSQOLKRVAVTPVKLAQIGTVV----- 122
DB 589 LPA--TVKVAASSPVVWNSPAT-----RMKLTAA-----AQVGTSVSATNTS 628
QY 123 ----VTVPRKSSVQSAVPTSVTVTGKPLNTVTLK-PSLS-GASSTPSEPNL----- 173
DB 629 TRPITVHRSQTV-IVAQAQAVTVTVGVTITLTVKSPISVPGSALISNLGKVMSTV 687
QY 174 --KAENSAVAQINLSPTMLENKCKKNFL--AMLILKACSGSOSP-----EMQGNVK 221
DB 688 QTKPVQTSANVTOASGTPYTOIOTKGPLPACTIILKLVTSADAGKPTTTITTTQASAGAK 747
QY 222 KLVQQLDAKLEAEFTRLLYVELKSSQPHLVPLKLSVALROLPLNSOSFIOOCVQO 281

```


Db 61 MLVSPQIVTRAEITSNITSRPAVPANPOTWIKICIVPNSSSOLIKKVAATPVKKLAQIGT 120
QY 121 TVTTPKPSVQSVAVPTSVVTPGKPLNTVTTLKPSLSAGSTPSNEPNKAENSA 180
Db 121 TVTTPKPSVQSVAVPTSVVTPGKPLNTVTTLKPSLSAGSTPSNEPNKAENSA 180
QY 181 VOINISPTMLENVKCKNFKLAMIKLACSGSOSPENGONVKKLVEOLDLAKIEAEFTTR 240
Db 181 VOINISPTMLENVKCKNFKLAMIKLACSGSOSPENGONVKKLVEOLDLAKIEAEFTTR 240
QY 241 LVEELKSSPOPHLVPFLKSSVVALROLPLNSOSFIOOCVOQTSDDNVIACTTTVTTSVP 300
Db 241 LVEELKSSPOPHLVPFLKSSVVALROLPLNSOSFIOOCVOQTSDDNVIACTTTVTTSVP 300
QY 301 VTTTSSSSSEKSIISGATAPRTVSQTLNPLAGVGAAGVTLHSVGTAAATGTTA 360
Db 301 VTTTSSSSSEKSIISGATAPRTVSQTLNPLAGVGAAGVTLHSVGTAAATGTTA 360
QY 361 GTGLQTSKPLVTSVANVTTVVSLQPEKPVSGTAVTSLPAVTEGTSGAALCLPSVAP 420
Db 361 GTGLQTSKPLVTSVANVTTVVSLQPEKPVSGTAVTSLPAVTEGTSGAALCLPSVAP 420
QY 421 VVSECDHICKPVIPTGVOIKLAQPGPVLSQP 452
Db 421 VVSECDHICKPVIPTGVOIKLAQPGPVLSQP 452

RESULT 2

US-08-188-582-16
Sequence 16, Application US/08188582

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comal, Lucio
APPLICANT: Dynlect, Brian D.
APPLICANT: Hoej, Timothy
APPLICANT: Rupert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-188-582-16

Query Match 23.2%; Score 521; DB 1; Length 737;

Best Local Similarity 34.0%; Pred. No. 2, 2e-36;
Matches 159; Conservative 61; Mismatches 128; Indels 120; Gaps 20;

QY 13 PPKVSSG-----PRLPAQIVAAKAPNTTITQEPANILQPLPGTVLKSNGLM 61
Db 68 PRATTSIGRATLTPVIALRPLPQ-----PQNPINIQ---NFQLPGMVLVSENGOLL 118
QY 62 LVSPOQTIVR-----AETSNITSRPAVPANPOTWIKICIVPNSSSOLIKKVAATPVKKL 115
Db 119 MI-PQALMOMQAOHAQOQTMAAPRPTSAAPVQISTVQAGPPIAR-QVTP----- 172
QY 116 AQIGTVVTTVPKPSVQSVAVPTSVVTPGKPLNTVT--TLKPSLSAGS-----TPS 168
Db 173 ----TTIIQV---SQOQTVQPSATLQSPGVQPOLVGGAAQVTSLSGTATAVQIGTQ 225
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFKLAMIKLACSGSOSPENGONVKKLVEOL 227
Db 226 RVPGATTTSSATE-----TWENYKCKNFKLSTLIKLSAGSOSTETANVKKELYOML 279
QY 228 LDKIAEFTKRLVYELKSSPOPHLVPFLKSSVVALROLPLNSOSFIOOCVOQ-----TS 283
Db 280 LDKIAEFTKRLVYELKSSPOPHLVPFLKSSVVALROLPLNSOSFIOOCVOQPPPTS 339
QY 284 SDNVIACTTIVTSPVTTTSSSSEKSIISGATAPRTVSQTLNPLAGVGAAGV 343
Db 340 Q-----ATTALTAVLVSSVQRTGKTAAYTSLQPPVLSL----- 376
QY 344 VTLHSVPTAAAGTGTAGLQTSKPLVTSVANVTTVVSLQPEKPVSGTAVTSLPAV 403
Db 377 -----TQPTQVGKQGOQPTLVIO-----QPPKP-----GALIRPQV 410
QY 404 TREGTSGAALCLPSVAPVVSFCDHICKPVIPTGVOIKL--AQPGVYL 449
Db 411 TLTQT-----PMVALRQPH-NRIMLTTPQDIOQLNPLOPVV 446

RESULT 3

US-08-646-715-16
Sequence 16, Application US/08646715

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comal, Lucio
APPLICANT: Dynlect, Brian D.
APPLICANT: Hoej, Timothy
APPLICANT: Rupert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/188,582

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 11.6146 seconds
(Without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452
Perfect score: 2249
Sequence: 1 GTLVTKVAPVSAAPKVS... VICTPVQIKLAPGVLSP 452

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: Issued_Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCPUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2249	100.0	801	1	US-08-725-012-2
2	521	23.2	737	1	US-08-188-582-16
3	521	23.2	737	1	US-08-646-715-16
4	241.5	10.7	921	1	US-08-188-582-2
5	241.5	10.7	921	1	US-08-646-715-2
6	204	9.1	2035	1	US-08-046-585-5
7	204	9.1	2035	1	US-08-393-703-5
8	204	9.1	2035	5	PCT-US93-11721-5
9	176.5	7.8	2972	4	US-09-579-181-2
10	176.5	7.8	3118	4	US-09-579-181-1
11	172	7.6	750	4	US-09-165-239A-4
12	171.5	7.6	1721	3	US-08-700-651-5
13	171.5	7.6	1721	3	US-08-928-361B-6
14	164.5	7.3	752	1	US-08-244-189-2
15	163.5	7.3	862	1	US-08-325-267A-4
16	161.5	7.2	1837	3	US-08-928-361B-5
17	159.5	7.1	894	3	US-08-362-525-22
18	159.5	7.1	894	3	US-08-971-692-15
19	156	6.9	941	4	US-07-757-022B-14
20	156	6.9	1022	4	US-07-757-022B-84
21	156	6.9	1038	4	US-07-757-022B-74
22	156	6.9	1049	4	US-07-757-022B-58
23	156	6.9	1140	4	US-07-757-022B-104
24	156	6.9	1270	4	US-07-757-022B-44
25	156	6.9	1311	4	US-07-757-022B-42
26	156	6.9	1313	4	US-07-757-022B-142
27	156	6.9	1314	4	US-07-757-022B-50

28	156	6.9	1320	4	US-07-757-022B-46	Sequence 46, Appl
29	156	6.9	1320	4	US-07-757-022B-60	Sequence 60, Appl
30	156	6.9	1354	4	US-07-757-022B-48	Sequence 48, Appl
31	156	6.9	1361	4	US-07-757-022B-40	Sequence 40, Appl
32	156	6.9	1363	4	US-07-757-022B-52	Sequence 52, Appl
33	156	6.9	1404	4	US-07-757-022B-2	Sequence 2, Appl
34	156	6.9	1404	4	US-07-757-022B-62	Sequence 62, Appl
35	155	6.9	907	3	US-08-783-774-2	Sequence 2, Appl
36	155	6.9	907	5	US-09-328-559A-1	Sequence 1, Appl
37	155	6.9	907	5	PCT-US95-04611A-19	Sequence 19, Appl
38	153	6.8	878	4	US-09-556-706B-2	Sequence 2, Appl
39	150	6.7	1537	1	US-08-325-267A-2	Sequence 2, Appl
40	148.5	6.6	805	4	US-09-103-429A-4	Sequence 4, Appl
41	148	6.6	786	4	US-09-103-429A-3	Sequence 3, Appl
42	146.5	6.5	3969	4	US-08-061-376-5	Sequence 5, Appl
43	142.5	6.3	2476	2	US-08-276-967-2	Sequence 2, Appl
44	140	6.2	1147	1	US-08-131-365B-38	Sequence 38, Appl
45	140	6.2	1147	2	US-08-668-123-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-08-725-012-2
Sequence 2, Application US/08725012
Patent No. 5710025
GENERAL INFORMATION:
APPLICANT: Dikstein, Rivka
APPLICANT: Tjian, Robert
TITLE OF INVENTION: B-cell Specific Transcription Factor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-012-2
Query Match 100.0%; Score 2249; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.9e-185;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GTLVTKVAPVSAAPKVS...VICTPVQIKLAPGVLSP 452
DB 1 GTLVTKVAPVSAAPKVS...VICTPVQIKLAPGVLSP 452
OY 61 MLVSPQGVTAETSNITSPAVPANPQYKICTVPSSSQLKKAVTVYKLAQIGT 120

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 31.7465 Seconds
(without alignments)
2933.659 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_452

Sequence: 1 GTLVTKVAPVSAPPKVSSGP.....VIGTPVQIKLAQPGPVLSP 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : SPTREMBL_21:*

```

1:  sp_archaea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_ornamentele:*
9:  sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	526	23.4	662	11	Q91W6	Q91W6 mus musculus
2	220	9.8	851	5	Q8T9E0	Q8T9E0 drosophila
3	215	9.6	1275	5	Q76602	Q76602 caenorhabdi
4	215	9.6	1349	4	Q8WMO4	Q8WMO4 homo sapien
5	209.5	9.3	528	6	Q28071	Q28071 sus scrofa
6	205	9.1	1979	11	Q9G935	Q9G935 mus musculus
7	202	9.0	1795	5	Q76894	Q76894 drosophila
8	200	8.9	245	11	Q61191	Q61191 mus musculus
9	198.5	8.9	842	5	Q9VGC9	Q9VGC9 drosophila
10	199.5	8.9	864	5	Q95U45	Q95U45 drosophila
11	199.5	8.9	864	5	Q9VGC8	Q9VGC8 drosophila
12	198	8.8	513	4	Q43418	Q43418 homo sapien
13	198	8.8	2045	11	Q9QWH2	Q9QWH2 mus musculus
14	197.5	8.8	1322	4	Q8T9E0	Q8T9E0 homo sapien
15	195	8.7	1079	5	Q9NA57	Q9NA57 caenorhabdi
16	192	8.5	1844	5	Q22579	Q22579 caenorhabdi

17	191.5	8.5	1893	5	Q9NKC9	Q9pkc9 drosophila
18	191	8.5	1286	4	Q15052	Q15052 homo sapien
19	191	8.5	2382	4	Q9H4A3	Q9H4a3 homo sapien
20	187	8.3	1029	4	Q8TDH7	Q8tdh7 homo sapien
21	186.5	8.3	873	10	Q9C548	Q9c548 arabidopsis1
22	186.5	8.3	5374	11	Q99ND0	Q99nd0 mus musculi
23	186	8.3	389	12	Q8V0N0	Q8v0n0 equine herp
24	185.5	8.2	3570	4	Q99552	Q99552 homo sapien
25	185.5	8.2	3971	3	Q96W66	Q96w66 schizosacch
26	185	8.2	648	4	Q14760	Q14760 homo sapien
27	183.5	8.2	2232	5	P91365	P91365 caenorhabdi
28	183	8.1	800	3	Q8TFG4	Q8tf94 schizosacch
29	182.5	8.1	374	12	Q8V0L6	Q8v0l6 equine herp
30	182.5	8.1	2187	11	P70670	P70670 mus musculi
31	181.5	8.1	356	12	Q8V0L7	Q8v0l7 equine herp
32	181.5	8.1	709	4	Q9NVJ9	Q9nvj9 homo sapien
33	181.5	8.1	961	3	Q92223	Q92223 emericella
34	179.5	8.0	645	4	Q9H048	Q9h048 homo sapien
35	179.5	8.0	867	12	Q93782	Q93782 equine herp
36	179.5	8.0	1324	4	Q15312	Q15312 homo sapien
37	178.5	7.9	622	4	Q14881	Q14881 homo sapien
38	178.5	7.9	1716	11	Q9JK31	Q9jkc1 mus musculi
39	178.5	7.9	1301	5	Q8TK34	Q8tk34 drosophila
40	178	7.9	826	12	Q8V0L5	Q8v0l5 equine herp
41	178	7.9	886	12	Q9Q8R7	Q9q8r7 human herpe
42	177.5	7.9	662	5	Q9VG00	Q9vg00 drosophila
43	177.5	7.9	3507	5	Q23587	Q23587 caenorhabdi
44	177	7.9	1236	3	Q9C105	Q9c105 schizosacch
45	176.5	7.8	1203	5	Q9NSK0	Q9nsk0 caenorhabdi

ALIGNMENTS

```

RESULT 1
O91MW6 PRELIMINARY; PRT; 662 AA.
ID O91MW6 AC O91MW6
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE TARA-binding protein associated factor TAFII135 (fragment).
GN TAF4A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-PO BRAIN;
RA Metsis M., Brunkhorst A., Neuman T.;
RT "Cell Type Specific Expression of the TrfID Component TAFII135 in the
RT Nervous System.";
RL Exp. Cell Res. 0:0-0(2001).
DR EMBL; AY038601; AAA94779.1; -.
DR MGD; MG1:2152346; Taf4a.
FT NON_TER 1
SQ SEQUENCE 662 AA; 71398 MW; 91A75F38CB0D0DA4 CXC64;

Query Match 23.4%; Score 526; DB 11; Length 662;
Best Local Similarity 35.0%; Pred. No. 3.5e-25;
Matches 155; Conservative 60; Mismatches 128; Indels 100; Gaps 18;

OY 20 PLPLPAQIVAAVAPNTTITTPANLPOLDPGVLIKSNSGPMLVSPQOTV-----TRAE 73
Db 10 PLPLPP-- --OKN- NIQ--NFQLPGMVLVSSENGQLMTI-PQAALMOAHQAQO 59
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 74 TTSNTISRPANPANTQTAKICTVPSSQSLLIKKVAVFEPVKLIAIGTVTTVPKPSVO 133
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 .PGSTMAPRPNPTGCAFPVQISTVAQNAGPIILAR-QVTPE-----TTIIKYV---SQAQ 107
      : ||| : :|| : : ||| : : ||| : : ||| : : ||| : : ||| :
OY 134 SVAVFSTVVVYTPGKFLMTVT--TLKPSSLGASS-----TPSNE-PNLKAENSAAVQINTL 185
      : ||| : :|| : : ||| : : ||| : : ||| : : ||| :

```

Db 108 TVVOPPTTLQSRSPGVQPOLVLGSAQASISGATATAVGTGTPORTVPGASTTSTATE---- 164
Oy 166 SPTMLENKKCNFLAMLIKACSGSQSPKMGONKVLLEOLDLAKTEMEETRKLYVEL 245
Db 165 ---TMBNKKKCSFLSTLLIKSSGQSTETANVLDLQNLIDGTEAEDFTSRLEY 221
Oy 246 KSSPPHLPVFPKRSVALROLPLNPSQSFIOOCVOOTSDVATCTTPTVTSPTVTT 305
Db 222 NSSPPVLPVFLKRSIPALROLPLPSAATIOSQOQPP---ASQATTLALTAIVLSSV 277
Oy 306 SSSQSEKSIISGATAPRTVSQTLNPLAGPVAKAGVTLHVSFPPTAGTTACTGLL 365
Db 278 QRTAGKTASVTSALQPPVYSL-----TQPTQVGVGKO 310
Oy 366 QTSKPLVSVANTVTTSIQPEKPVVSGTAVTSLPVTFFGTSALCLPSVKPVVSTC 425
Db 311 ADPTPLVIO-----QPPKP-----GALIRPQVTLTQT-----PVAALR 344
Oy 426 WDICKPVIGTPVQIKLAPGPV 448
Db 345 QPH-NRIMLTTPQOIQLOLQOPV 366

RESULT 2

Q8T9EO PRELIMINARY: PRT: 851 AA.
AC Q8T9EO:
DT 01-JUN-2002 (TREMREL. 21, Created)
DT 01-JUN-2002 (TREMREL. 21, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE SD04735P.
GN TAF10.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stepleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Lao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RA Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY069807; AAL39952.1;
SO SEQUENCE 851 AA; 92093 MW; 665B28B9586C984C CRC64;

Query Match 9.8%; Score 220; DB 5; Length 851;
Best Local Similarity 26.5%; Pred. No. 7.7e-06;
Matches 117; Conservative 59; Mismatches 146; Indels 120; Gaps 22;

Oy 22 LPAPQIVAVK--APNTTITQ--FPANLQLPPTGVLIKNSGPMPLVSPQOTVRAETTSNI 78
Db 93 LPAGVVGKROQAPSOQOQKNMPTN---PLSRVYINSHAGVROPSP--SMTNTTATSN 148
Oy 79 TSPRAVPANPOTVVKICTVPSNSS--QLIKKVAVTVPYKKAQIGTVVTVTPKPSVQSA 136
Db 149 I-----VNSVASSGTANSSQPHLTOINAOP--QLPOL--TQIQITPAQOQSOQ--- 193
Oy 137 VPTSVVTVTPGKPLNTVTTLKPSLSGASTSPNEPNLKAENSAVAOINLPTMLENVK-K 195
Db 194 -----QOVNNSVAGSTATAVASTTA-----ATTTOQGNTRK 226
Oy 196 CKNFLAMLIKACSGSQSPKMGONKVLLEOLDLAKTEMEETRKLYVELKSSPPHLPV 255
Db 227 CRKFLANIEL--STRERKPYEKNVRTILOELVNNANVEEFCRDLRLNLNPSQCLIG 284
Oy 256 FLKKSVALROL-----LPSQSFIO-----QCVQO 281
Db 285 FLKKSPLRLKQALYKELVIEGIRPPQHVIGLAGISQOLPRLQAIQIRIGSPQTTTIOQ 344
Oy 282 TSSDVAICTTCTTPTVTSPTVTSVSSQSEKSIISGATAPRTVS---VQTLNPLAGPVG 338

Db 345 TOYRMI--TPNALGTPRPITIGHTTISKOPPN---IKLPAPRLVNTGIGTQIP-SLOVP 398
Oy 339 AKAGVTLHVSFGPTAATGTTACTGLQTSKPLVTSVANTVTTSIQPEKPVVSGTAVL 398
Db 399 GQANIYQIR--GPQHQLOLQRTSGVSQIRATTRP-----PNSVPTAN----- 436
Oy 399 SLPAVTFGTSAGALCLPSVKP 420
Db 437 KLVAVKVGQTLKAI--TPSLHP 457

RESULT 3

Q76602 PRELIMINARY: PRT: 1275 AA.
AC Q76602:
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE Hypothetical 122.9 kDa protein.
GN H02F09.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
OC Rhabdilitae; Peloderinae; Caenorhabdilitis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berke M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Faveille A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Smaildon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RN Nature 368:32-38(1994).

[2]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RC Geisel C., Harmon G.;
RT "The sequence of C. elegans cosmid H02F09.";
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.

[3]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RC Waterston R.;
RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF07538; AAC64622.1;
KW Hypothetical protein.
SO SEQUENCE 1275 AA; 122924 MW; CB59B70C05959E25 CRC64;

Query Match 9.6%; Score 215; DB 5; Length 1275;
Best Local Similarity 23.6%; Pred. No. 2.6e-05;
Matches 110; Conservative 71; Mismatches 182; Indels 104; Gaps 18;

Oy 2 TLTVKAPVAPSPKXSSGF-----RLPAPQIVAKAPNTTITQFPANLQLPPTGVLIKSN- 56
Db 275 TVTVPTTVTGPTTVTVTVTVTVTVTVTSPTTPSPSTVTVTVTVTVTVTVTVTVTKPST 334
Oy 57 --SGPLMVS--POQVTVRAETTSNITSRAVPANPOTVICTVPSNSSQLIKKVAVTPK 113
Db 335 VVTABSTVTVSTVTVTKNTV--VTSSPTVATTPTV--VTPPT-----VTVPS 382
Oy 114 KLAQIGTTVV-----TVVPPSSVQSAVAPTSTV-----VTPGKPLNTV 153
Db 383 TVTVPTTV 440

QY 154 TILKPSLSGASTPSNPNKAENSAVOILSPMLLENKCKNFLMLIKLACSGS 213
DB 441 AVTTETISGSSSTP-----LPSOSTSLMSLSSTYPS-----STAGATS 481
QY 214 PEMGONKVKLEQLLDA--KIEAEFPRKLYVELKSSPOHLVPIKKSVALQLLPNS 271
DB 482 PATQOQKRPITIGTSMSSGPTTAPGASTESTVLOSSTPSGTTV-----TLPSG 529
QY 272 QSPFIQOCVOOQSDMVIATCTTCTTTPSVTTTSSSQSEKSIIVSGATAPRTVSOTLN 331
DB 530 SSTAGTSPQASVTVTYTDISTVSGSVTISQTAESSLSTSPSAGSSI--STVSTVSQ 588
QY 332 P-----LAGPVGAKAGVYTLHSGVPPAATGGTATGILLQTSKPLVTSVA 376
DB 589 PSTYIPVSSASISYTLSTSGSTGSTA-----SPGTSSSGSTSGSPSTISGSS-----A 636
QY 377 NTVTYVLSLOPEKPVSGTAATVLSLPAYTFGTSGAATCLPVRKPVYS 423
DB 637 STVYG-STVTEASTISGTESTSTIPGSTESTVSEASTVSGSVSTVS 682

RESULT 4

08MMO4 PRELIMINARY; PRT: 1349 AA.
AC Q8MMO4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Mucin 5 (Fragment).
GN MUC5AC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21426417; PubMed=11535137;
RA Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
RT repetitive region.";
RL Biochem. J. 358:763-772(2001).
DR EMBL: AJ298318; CAC83675.1; .
FT NON_TER 1
FT NON_TER 1349
SQ SEQUENCE 1349 AA; 135600 MW; 4DC3C1544F1E5EBA CRC64;

Query Match 9.68; Score 215; DB 4; Length 1349;

Best Local Similarity 23.28; Pred. No. 2.8e-05; Mismatches 216; Indels 110; Gaps 20;

QY 2 TLVTKVAP---VSAPPKVSSGP---RLAPQIYAVKAPNTTTOFP-ANLQLPQCVLI 53
DB 725 TSTSTAPTASTASPTSTSSAPPTNTNTSAPTSTTSAPISTISAPTSTTSTPQTSTI 784
QY 54 KNSNGPLMLVSP-OQYVTRAETTNITSRAVPANPOTVKTICVYVNSSOLIKKAVATPV 112
DB 785 ---SSPTTSTTPOTSTSTSTSTAP-----TSTTSAPTST-----TSTPQ 828
QY 113 KKLAIQIGTVVTPKPSV-----QSVANPTSVVTVTPCKPLNTVTTLKPS 159
DB 829 TSTISAPTSTSTAPVTAISTISAPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 888
QY 160 SLGASSTPSNEPNL-----KAENSAVAQIULSP----- 187
DB 889 S-GSGTTPSPVPTSTASVSKTSHSVSKTTHSQPTRYRCHRCWTKMFVDWDFSPG 947
QY 188 -----TMLENVKCKKFLMLIKLACSGSOSP-----MGONVKKVLEQLLDKTI 232
DB 948 PHGDKETNNYNIIRSGEIKRPREIETRIACRASHPESVTEHGOVYVQCREGLVCRN 1007
QY 233 EAEFTKRL-----VELKSSPOHLVPIKKSVALQLLPNSOFLQOCVOOQSDMVI 288
DB 1008 QDOQGFPRKMLNTEYVLCCTETPKGCPVTSTPVA-----PSTPSGRATSTPSTSSMOK 1062

QY 289 ATCTTVTTSBPVTTTSSSQSEKSIIVSGATAPRTVSOTLNPLAGPVGAKAGVYTLHS 348
DB 1063 SRTTLVTTSTSTPQSTSTAPSTSTIPAST-PSTTAPPTSTSTAPPTSTSTAPHT 1121
QY 349 V-GPI-----AATGTTAGTGLQTSKPLVTSV-ANVTYVLSLOPEKPVSGTAATVLSLP 401
DB 1122 TSGPTSTTLAPTSTSTAPPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTP 1181
QY 402 -----AVTFEGTSGAATCLPVRKPVYS 423
DB 1182 QTSKTSATSTSTSGC-TPSPVPTTS 1208

RESULT 5

ID Q29071 PRELIMINARY; PRT: 528 AA.
AC Q29071;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Gastric mucin (Fragment).
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=95275264; PubMed=7753593;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin.";
RL Biochem. J. 308:89-96(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lamont J.T.;
RT "Pig gastric mucin: isolation and characterization of a cDNA clone
RT with a novel tandem repeat.";
RL Gastroenterology 106:200-200(1994).
DR EMBL: U10281; AAC48526.1; .
FT NON_TER 1
FT NON_TER 528
SQ SEQUENCE 528 AA; 49907 MW; 0BF0F6879203B2EA CRC64;

Query Match 9.38; Score 209.5; DB 6; Length 528;

Best Local Similarity 23.98; Pred. No. 2e-05; Mismatches 153; Indels 137; Gaps 24;

QY 8 APVSAAPPV---SSGRLAPQIYAVK-----APNT--TTIOFPANLQLP--GTVLIK 55
DB 108 APTSTSTVOPSSGVSPTSTATSVQSSSSSAPTSTATSVQ-PSSSSSPPTSTVSVQ 166
QY 56 NSGPLMLVSPQOYVTRAETTNITSRAVPANPQVKTICVYVNSSOLIKKAVATPVK- 113
DB 167 SSS-----SAPTSTATSVQPSSSSPPTSTVSVQTSSSSVPTSTSTVOPS 215
QY 114 KLAQIGTVVTV-----PKSSVQ-----SVAPTSVVTVTPCK-----PLNTVTT 155
DB 216 SSSSVPTTSTATSVRSSSSSPTIPPTSTVQPSSSSAPTSTATSVQPSSSSPTIPPTST 275
QY 156 LKPSLSGASSTP-----SNEPLKAENSAVAQIULSPMLLENKCKKFLMLIKLACSG 210
DB 276 VOPSS---SSAPPTSTATSVQPSSSSPTSTSTISVQPS-----SS 313
QY 211 SOSPEMGONVKKVLEQLLDKTIKIEAEFPRKLYVELKSSPOHLVPIKKSVALQLLPN 270
DB 314 SSSPTTS-----TSTVOPSSSGSAPTSTATSVQPSSSSPPPI-----SSTISVQ---PS 359

```

OY 271 SOS-----FIQ-----QCVOQTSDDVIACTTNTS-----PVYTTTV 305
DB 360 SSSSSSTTSTTSVOPSSSGSAPTTSATSVOPSSSSSVPTTSATSVSSSSSPPIPTTTS 419
OY 306 SSSOSEKSIIVSGATAPRTVSVOTLNPLAGPVAKAGVTLHSGVTAATG--TIAAGTGL 364
DB 420 VOPSSSSSVPTTSAT-----SVOTSSSSSPPIPTST-----SVOPSSSSSATTATATSV 468
OY 365 LOTS-----KPLVTSVANTVTTVSLOPEKPVVSGIATVLSLPAVTFTGSGAA 412
DB 469 QPSSSSSPPISTTSVOPSSSSPTTSTTSVOPSS-----SGSAPTTSATSVOPSSSSSP 525
OY 413 I 413
DB 526 I 526

RESULT 6
OYQ35 ID 090Y35 PRELIMINARY; PRT; 1979 AA.
AC 090Y35: 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Host cell factor C1 (Fragment).
GN ABCD1 OR HCFC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RA Platzner M., Brenner V., Reichwald K., Wiehe T., Oksche A.,
RA Rosenlhal A.;
RT "Comparative sequence analysis of the mouse Licam locus and the
RT corresponding region of human Xq28."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF13093; AAF2156.1;
DR MGD; MGI:105942; Hcfcl.
DR MGD; MGI:1349215; Abcd1.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; kelch; 4.
DR SMART; SM00060; FN3; 1.
FT NON-TER 1
SQ SEQUENCE 1979 AA; 203436 MW; ABE998C3DDE688A9 CRC64;

Query Match 9.1%; Score 205; DB 11; Length 1979;
Best Local Similarity 23.7%; Pred. No. 0.00019;
Matches 128; Conservative 63; Mismatches 209; Indels 140; Gaps 26;

OY 4 VTKVAVSAPKPVSSGPRP---PAPQIVAVKAPNTTIOFPANL-----QLPPTGTVL-- 52
DB 417 VPAVLKVTGPQATGTPPLVTMPRASQ--AGKAPVTVT-SLPSAVRWVVPQSGVQGVIGS 473
OY 53 -----IKNSGPLMVSPQGYTTRAEITSNITSRAVPANPQTVKI 93
DB 474 NPQMSGMAALAAAAAATQKIPSSAPTVLSVPAGIT-----IVKTVAVTGGTTLLPATVRY 529
OY 94 CTVP-----NSSQLIKKVVV---TPVKLAQIGTIVTVTPKSSVQSVAVPISVTVTP 146
DB 530 ASSPWWVSAPATRMKLTAAAOQVTSVSSAANSTRITIVHKSQV-TVAQQAQVTVTVV 588
OY 147 GRPLNTVTTLK-PSLS-GASSTPSNEPNL-----KAENSAVQVIMLSPMLENVKCKN 198
DB 589 GGVTKTITLVKSPISVPGSALISNIGKVMVSVQTPQTSANTGASGVPVQIIQIFNG 648
OY 199 FL--AMLTILASGSQSP-----EMQONKYLVEQLDAKTEAEFTTKLYVELKSS 248
DB 649 PLPAGTITLVKSADGKPTTITTTQASGAGTKPTILGI-----SS 689
OY 249 PGPHLVPLPKKRSVVALRQLPNSQSFIOQVOQTSSDWMVIACTTTVT-----S 298

```

```

DB 690 VSPSTT---KPGTTTTIKTIPMSAIIITQAGATGVTSSPGIKSPITITTKVMTSGTCAPA 746
OY 299 PVVT-----TVSSOSEKSIIVSGA-----TAPRT-----VSQVTLNPLAGPV 337
DB 747 KITPAVKIKTHGQGCQVNOVLKAGAPGCPQRTILRTVPMKGVALLYMPVYSAVKPAVTLT 806
OY 338 GAK--ACGVTLHSGVPLAATGCTTACTGLLQTSKPLVTSVA--NVVTVSLOPEKPVSG 393
DB 807 VVKGTGTVTLGIVTGIVST--SLAGAGHSTASLALPTITLGTATLSSQ-----VINP 860
OY 394 TAVTLSPAVTFEGTSGAATCLPSVKPVYSCFMDHICKPVGIVQOKL-AQGQPVLSQP 432
DB 861 TAITVSAQTTLLTAAGGLTPTTTPMPV-----SQPTQVTLITAPSGVEAOP 907

RESULT 7
OYQ35 ID 076894 PRELIMINARY; PRT; 1795 AA.
AC 076894: 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE EG:56G7.1 protein.
GN EG:56G7.1 OR CG14796.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Flanck C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Broksstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cusley S., Dahlke C., Davaport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapran G.H., Ke Z., Kenison J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Put V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spires E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasthman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185(2000).

```


RP SEQUENCE FROM N.A.
 RA Cadieu E., Dreano S., Lelaure V., Motlier S., Galibert F.;
 RT "Sequencing the distal x chromosome of Drosophila melanogaster."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Benos P.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003421; AAF4564.1; -
 DR EMBL; AL031028; CA19845.2; -
 DR FlyBase; FBgn0025390; EG:5667.1.
 DR InterPro; IPR002555; Chitin_bind_Pera.
 DR InterPro; IPR002965; P-rich_extensn.
 DR Pfam; PF01607; CBM_14; 2.
 DR PRINTS; PRO1217; PRICHEXTENSN.
 DR SMART; SM00494; ChIBD2; 2.
 SO SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 9.08; Score 202; DB 5; Length 1795;
 Best Local Similarity 21.88; Pred. No. 0.00026;
 Matches 95; Conservative 58; Mismatches 196; Indels 86; Gaps 13;

QY 9 PVSAAPKVSQPRLPAPQIVAVKAPNTTIOFANLQ-----PGTVLKNSSG 58
 Db 700 PFSSTGPTTPK-PSRTTPTTKVITTOITTPPLKRSSTTSQTPPTTPQPTT 758
 QY 59 PLMLVSPQOVTVAETTSNITSRPVAPANDQVVKICTVPNSSSOLIKKVAATPVKLAQI 118
 Db 759 TTLVTFKTSVTTTTEKPISSPK-PTTTQKTSTAPNNT-----KVATTOKEPTPT 812
 QY 119 GTT-----VVTVPKPS-----VQSAVAPTSVTVTPGKPLNT 152
 Db 813 QSTSTFTTRKTTNNPEPTSTKPTSTTPKPSVTSVASTKSTISSPKPTE 872
 QY 153 VTTLPSSSLASSPSPNEPNKKAENSAVOQINISPTMLENKKKKFLAMIKIACGSG 212
 Db 873 KSTNPSTNSYKISALSSSTORA-----TSTSEPTKTONTTTTTPKPTTKTS 922
 QY 213 SPENGWKKLVLEOLDLAK--IEAEFTKRLVELKSSPOHLPFLKSSVALROLIPN 270
 Db 923 TQKATTSQKSVYITTKKATESPLTTLSTEBNTPK-----LRTTTP 970
 QY 271 SOSTQOCVOQTSNDVIATCTTTVTTSPTVTTVSSQSEKSIIVSGA--TAPR--TVS 326
 Db 971 TTS-----VTATRTITTTTISSTSTSTQKPSSTPTSTRTPTKVTVI 1017
 QY 327 VQTLNPLAPGAKAGVATLHSGVPAATGTFAGTGLQTSKPLVSVANTVTVSLQ 386
 Db 1018 VSTONPTT--TTSKTSVTITTPNPSPSTORPTTTTQPTSTASTSIGTRIPPTTP 1075
 QY 387 EKPVSCTAVTSLP 401
 Db 1076 QNSTSTDLTTPTRP 1090

RESULT 8
 Q61191 PRELIMINARY; PRT; 2045 AA.
 AC 061191;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Transcription factor C1 (HCF).
 GN HCF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TTSUE-LIVER;
 RA Kistle T.M., Dasher R.;
 RT "CDNs encoding the mouse homolog of the human transcription factor C1

RT (HCF).";
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53925; AAB01163.1; -
 DR MGD; MGI:105942; Hcf1.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR001796; Kelch.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF01344; Kelch; 5.
 DR SMART; SM00060; FN3; 1.
 SO SEQUENCE 2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

Query Match 8.98; Score 200; DB 11; Length 2045;
 Best Local Similarity 24.18; Pred. No. 0.0004;
 Matches 129; Conservative 68; Mismatches 201; Indels 138; Gaps 28;

QY 7 VAPVAPKVSQPRLPAPQIVAVKAPNTTIOFANLQPPGVILKNSGPMV 63
 Db 486 VLKVTGPQATGTGTLVTRPQSO--AGKAPTVV-SLPASVPM--VPTQSAQGTIVGS 539
 QY 64 SPQQT-----VTRAETTSNI--TSRP--AVPAN-----PQVKTCTVP 97
 Db 540 NPQMSGMAALAAATAQKIPSSAPATAMSVPACTTVKTVAVTPGTTLPATVAVASSP 599
 QY 98 ----NSSQLIKAVV--TPVKLAQIGTTVTVTPKPSVQSAVAPTSVTVTPGKPL 150
 Db 600 VMVSNPATRLMLKTAAGVGTSSAANTSRPITVHKSGTV--TVAQQAQVTVTVGCVT 658
 QY 151 NTVTTLK-PSL-GASSPSPNEPNL-----KAENSAVOQINISPTMLENKKKKFL-- 200
 Db 659 KTLTVKSPISIPGSGALISMLGKMSVQVOTKPVQTSVAVTQASTGPTVTOIQKGLPA 718
 QY 201 AMLIKIACSGOSP-----EMQWKKLVLEOLDLAKIEAEFTKRLVELKSSPOH 252
 Db 719 GTIKIVTSADCKPTTITTTQASGAGTKPIIGI-----SSVSPS 759
 QY 253 LVPLKKSVALROLPLNSOSFIQOCVOQTSNDVIATCTTTVT-----SPVPT 302
 Db 760 TT---KPGTTTIIKITPMSAITTQAGATGVSSPGIKSPIITITTKVTSGTGAPAKIT 816
 QY 303 ----TVSSQSEKSIIVSGA-----TAPRT-----SVQTLNPLAPGAK- 340
 Db 817 AVPKIATGGOQGVQVYVLAAGAPGPGTILKTVPMGVRVLTPTVAANVPAVTVLVKG 876
 QY 341 AGVTLHSGVPAATGTFAGTGLQTSKPLVSVV--MTVTVSLQPEKPVSGAVT 397
 Db 877 TTGVTTLGTVIGVST--SLAGAHSTASLAPITTLGIAITLSSQ-----VINPAIT 930
 QY 398 LSLPVAFTGTSAGALCLPSVKPVSFSCMDHCKPVIQTVQIKL-AQPGVLSQ 452
 Db 931 VSAQOTLTLAAGLTTPPTITWQPV-----SQPTQVTLTAPSGVEAQP 973

RESULT 9
 Q9VGC9 PRELIMINARY; PRT; 842 AA.
 AC 09VGC9;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CPN protein.
 GN CPN OR CG4795.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Ephygrotia; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephygrotidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
 RA Abri J.F., Agbayant A., An H.-J., Andrews-Plankoch C., Baldwin D.,
 RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu T., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacheb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003694; AAF54755.1; -
 DR Flybase: FBgn0010218; Cpn.
 DR InterPro: IPR002965; P. rich. extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 SQ SEQUENCE 842 AA; 82242 MW; D71E531327EF8501 CRC64;

Query Match 8.9%; Score 199.5; DB 5; Length 842;
 Best Local Similarity 22.5%; Pred. No. 0.00015;
 Matches 124; Conservative 70; Mismatches 203; Indels 153; Gaps 25;

QY 1 GTLVTV-APVSA-----PVSSGPRLPAP---QIVAKAPNTTIOFPANIQ 45
 DB 4 GTIPSEVSAFVAPVPSAFAVQVSPAAVAPAPAPAPAPAPAPAPAPAPAPAP 63
 QY 46 LPPGTVLINSNGPLMLVSPQOTVTRAEFTTSNTTSRPAPAPAPAPAPAPAPAP 104
 DB 64 IPAPADIAAASVAPVASVAP--PVVAAPTPP-----NASVSTPPVAVAQIPVAVSAPVA 116
 QY 105 KRYAVT-----PVKKLAQIGTT--VVTIVKPSVSVANPTSVVTVTGKPLNTYTLK 157
 DB 117 PVAATPTTPAPAPVAPVAPVATPVAASAPTPAAVTPVSP--VIAITPPVANTIV--- 171
 QY 158 PSSLGASSTPSPNEPNKAENSAVQINLSPTMLE-----NYKKCNFLAMLI 204
 DB 172 PVAAPVAAVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 228
 QY 205 KLASGSGSPENQONKVLIEQLDAKIEAEFTRLYELKSSPPHVLVPLFKKSVAL 264
 DB 229 -----PEVSVAATKPLAAAEPPVAPAPATETPVAPAAASPVSAVAPAVETAIVA- 278
 QY 265 ROLPNSGFIQOCVOQTSDMYIATCTTWTTSPTVTTTSSGSEKSIIV---SGAT 320
 DB 279 -----PVASAS-----TEPPVAAATLTTPAPET--PALAPVVAESQVAAANTVATPTTPAP 325
 QY 321 APRTVS-----VQTLNPLAGPYGAKAGVTTLSVGPFTAATGTTAGTGLL 365
 DB 326 EPEITAPVVAETPEVASVAVAETTPPVVAPVAES-----IPAVVAT----- 369

QY 366 QTSKPLVTSVANTVTVSLQPEKP-----VSGIAVTLIS-----LPAYT 404
 DB 370 -TPVPATIAVTDDBVTASAVBELPPIAPSPVSAVAETPEVIAAPVLPVAPAEVPAVY 428
 QY 405 EGET-----SCAATCLDSVRYVSFCMD-----HICKPVIGTP-----VQIKL 442
 DB 429 AEETPETPAPASAPVTIAALDIEVAPVIAAPSDAPAPAEASAPVITPTTASVPETT 488
 QY 443 AOPGPVLSQP 452
 DB 489 APPAAVTEP 498

RESULT 10
 Q95045 PRELIMINARY; PRT; 864 AA.
 AC Q95045;
 DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-DEC-2001 (TEMBLrel. 19, last sequence update)
 DT 01-MAR-2002 (TEMBLrel. 20, last annotation update)
 DE GH08002P.
 GN CPN OR CG4795.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritidae; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayant A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacheb J., Paragas V., Park S., Phuananavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBI databases.
 DR EMBL: AY058315; AAL13544.1; -
 DR Flybase: FBgn0010218; Cpn.
 DR SEQUENCE 864 AA; 84565 MW; E2B96CFB0CB9802 CRC64;

Query Match 8.9%; Score 199.5; DB 5; Length 864;
 Best Local Similarity 22.5%; Pred. No. 0.00015;
 Matches 124; Conservative 70; Mismatches 203; Indels 153; Gaps 25;

QY 1 GTLVTV-APVSA-----PVSSGPRLPAP---QIVAKAPNTTIOFPANIQ 45
 DB 4 GTIPSEVSAFVAPVPSAFAVQVSPAAVAPAPAPAPAPAPAPAPAPAPAPAP 63
 QY 46 LPPGTVLINSNGPLMLVSPQOTVTRAEFTTSNTTSRPAPAPAPAPAPAPAPAP 104
 DB 64 IPAPADIAAASVAPVASVAP--PVVAAPTPP-----NASVSTPPVAVAQIPVAVSAPVA 116
 QY 105 KRYAVT-----PVKKLAQIGTT--VVTIVKPSVSVANPTSVVTVTGKPLNTYTLK 157
 DB 117 PVAATPTTPAPAPVAPVAPVATPVAASAPTPAAVTPVSP--VIAITPPVANTIV--- 171
 QY 158 PSSLGASSTPSPNEPNKAENSAVQINLSPTMLE-----NYKKCNFLAMLI 204
 DB 172 PVAAPVAAVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 228
 QY 205 KLASGSGSPENQONKVLIEQLDAKIEAEFTRLYELKSSPPHVLVPLFKKSVAL 264
 DB 229 -----PEVSVAATKPLAAAEPPVAPAPATETPVAPAAASPVSAVAPAVETAIVA- 278
 QY 265 ROLPNSGFIQOCVOQTSDMYIATCTTWTTSPTVTTTSSGSEKSIIV---SGAT 320
 DB 279 -----PVASAS-----TEPPVAAATLTTPAPET--PALAPVVAESQVAAANTVATPTTPAP 325
 QY 321 APRTVS-----VQTLNPLAGPYGAKAGVTTLSVGPFTAATGTTAGTGLL 365
 DB 326 EPEITAPVVAETPEVASVAVAETTPPVVAPVAES-----IPAVVAT----- 369
 QY 366 QTSKPLVTSVANTVTVSLQPEKP-----VSGIAVTLIS-----LPAYT 404

```

Db 370 -TPVPATLAVTDPDVTASAVDELPPVIAVSPVSAVAVETPDVLAPVILPPVAABEPVAVV
QY 405 FGET-----SGAALCLPSVKPVVSFCMD-----HICKPVIGTP-----VOIKL 442
Db 429 AEEIPERPAPASAVPTIALDIPVAVPIAASDAAPASAAPIVSTPTTASVPETT 488
QY 443 AOPGPVLSQP 452
Db 489 APPAAVPTPEP 498

RESULT 11
QYVC8 ID QYVC8 PRELIMINARY; PRT: 864 AA.
AC QYVC8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Cpn protein.
GN Cpn OR CG4795.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Adell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Bokoyva D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Buttle K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazuelos M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.C.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Welshlock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang M., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "the genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003694; AAF54754.1; -.
DR FLYBase: FBgn0010218; Cpn.
DR InterPro: IPR002965; P-rich extensn.
DR PRINTS: PR01217; PRICHEXTENSN.
SQ SEQUENCE 864 AA; 84595 MW; 5B3635A2835EC658 CRC64;

```

```

Query Match 8.9%; Score 199.5; DB 5; Length 864;
Best Local Similarity 22.5%; Pred. No. 0.00015;
Matches 124; Conservative 70; Mismatches 203; Indels 153; Gaps 25;

QY 1 GTLVTKV-APVSAV-----PKVSSGRLAP-----QIVAKAPNTTITPPANLQ 45
Db 4 GTIPSPVAPVAPVTPSAVAPVQVVSPPAAVAPAPAPVAVTPVAPPTLASVQPAVVT 63
QY 46 LPFGTVLKSNSGPMILVSPQGVTRAFETTSNTSRPVPANPQVTKCTVPSN-SSQLI 104
Db 64 IPAPAPVIAASVAPVAVSAP--PVVAAPTPP-----AASPSTPPVAVAQIPVAVSAPVA 116
QY 105 KKVAVT-----PVKLAQIGTT--VVTTPRPSVQSVAVPTSVVTPVGRPLNTVTLK 157
Db 117 PPVATPPVAPVAPVAPVAVIATPPVAPASAPVPAVTPVSP--VIAFPVVPANVTV--- 171
QY 158 PSSIGASTPSNEPRKAKNSAANOINLSPMLP-----NVKCKKFLAML 204
Db 172 PVAPVAVAVPAPVAVPAPVAVLAPVAPVAVAPVAVETPPVAPVAPVATIPPC---VAPLI 228
QY 205 KLAOSGSPMGONVKKVLEQLLDIAKEAEFFRKLVELKSPQPHLVPLKRSVAL 264
Db 229 -----PEVSVATKPLAAEBVVAVAPATETPVAPAAASHVAVAPAVETAVVA- 278
QY 265 RQLPNSQSFLOOCVOQTSSDMVATCTTVTTSPTVTTVSSQSEKSIIV---SGAT 320
Db 279 ----PVSA-----TEPPVAAATLTLTAPET--PALAPVAAEQVAVANTVATPPAP 325
QY 321 APRVVS-----VQTLNPLAGPVKAGAVVTLHSGVPIATGTTAGTGL 365
Db 326 EPETIAPVVAETPEVAVSAAVETPPVPPVAAES-----IPAVVAT----- 369
QY 366 QTSPIVTSVANTVTVVTSLOPEKP-----VSGTAVTIS-----LPAVT 404
Db 370 -TPVPATLAVTDPDVTASAVDELPPVIAVSPVSAVAVETPDVLAPVILPPVAABEPVAVV 428
QY 405 FGET-----SGAALCLPSVKPVVSFCMD-----HICKPVIGTP-----VOIKL 442
Db 429 AEEIPERPAPASAVPTIALDIPVAVPIAASDAAPASAAPIVSTPTTASVPETT 488
QY 443 AOPGPVLSQP 452
Db 489 APPAAVPTPEP 498

RESULT 12
QYVC8 ID QYVC8 PRELIMINARY; PRT: 513 AA.
AC QYVC8;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Intestinal mucin (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97476275; PubMed=9334251;
RA Gum J.R., Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,
RA Vinal L.E., Robertson A.M., Swallow D.M., Kim Y.S.;
RT "MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl
terminus, and a novel upstream repetitive region."
RL J. Biol. Chem. 272:26678-26686(1997).
DR EMBL: AF007190; AAC02268.1; -.
DR PRINTS: PR01608; BACINVASINC.
FT NON_TER 1
FT TER 513
SQ SEQUENCE 513 AA; 52573 MW; 3BBAC2AFAAE1436 CRC64;

```

Query Match 8.8%; Score 198; DB 4; Length 513;
 Best Local Similarity 20.8%; Pred. No. 0.0001;
 Matches 110; Conservative 81; Mismatches 195; Indels 142; Gaps 22;

10 VSAPPKVSQGRPLPAQIVAKAPNT-----TTIOFP--ANLQPL-----PGTV 51
 2 MTSPSPVSS-----SITPTNTMTSMRTTYMPTANTLISPLSSILSTPVSSTE 51
 52 LIRS---NSQPL--MLVSPQOIVTRAFTTSNIT--SRPAVPANQIVKI----- 93
 52 MISHNTNTPRLSTLVLTLLTITRSPTESETTPTSPSTIVSDSTELIVSTISITGLS 111
 94 --CIVPNSSQOL--IKKVAATPVKKLAQIGTIVTVTPK-----PSSVQSV--AVPTS 140
 112 TATLPTSSSLPTETATMPT-----TLLITTPNTSHSTPSFTSSTIYSTVTS 164
 141 VVTVPCKPLN---TWTLKPSLSLASSPSPNEP-----NLK 174
 165 TPAISSASPTSGTMVNTSTMTPSL--STDTPSTPTTITTPVSGTGLTATDPTSTF 223
 175 AENSAAVOILSPMLNKKCNFLMLIKLACSGSQSPKGVKVLQIDAKIEA 234
 224 VSSSSAKSKSVIPSS--PSIONTERSLVSMTSATTPSLRPTITSTDTLSLITPST 282
 235 EEFTRKLYE-----LKSSPQRLVPEFLKSVVALAQLP----- 269
 283 YSFSSSKSASACTTHTETISSLPASTNTHTTAESALATTTTSTSTPMEPSTVA 342
 270 -----NSQFIQOCVQOOTSMDVINTCTTVTTPSVVTVTVSSQSEKSI--IV 316
 343 TTGCGQTTFPSSTAFLETTTLPTTDFSTESLTLTAMTSPTTSSITSPIDIMSMHTT 402
 317 SGATPPTVSQTLNPLAGVGVAKAGVYTHSVGPATAGTATGTLQTSKPLVTSVA 376
 403 SMTATNTLSPILSSILSSPVPSTEVTTSTHT-----NTNPVSTLVTPP--ITTR 453
 377 NPTVTVSLOPEKPVSGTAVTLSP--AVTEGEGSAAICLPSKVPVVS 423
 454 STLTSEIAYSSPSTSTESTETITPTTMTETSTATSLPPTSSLVS 501

RESULT 13
 09QWH2 PRELIMINARY; PRT; 2045 AA.

AC 09QWH2; (TREMblrel. 13, Created)
 DT 01-JUN-2002 (TREMblrel. 13, last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
 DE C1 transcription factor.
 GN HCF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kristie T.M.;
 RT cDNAs encoding the mouse homolog of the human transcription factor C1
 (HCF).
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 080821; AAD09225.1;
 DR MGD; MGI:105942; Hcfcl.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF01344; Kelch; 5.
 DR SMART; SM00060; fn3; 1.
 SQ SEQUENCE 2045 AA; 210546 MW; 7AD38FCD78EABA9C CRC64;

Query Match 8.8%; Score 198; DB 11; Length 2045;
 Best Local Similarity 24.1%; Pred. No. 0.00054;
 Matches 129; Conservative 67; Mismatches 202; Indels 138; Gaps 28;

7 VAPVAPPKVSSQGRPL--PAQIVAKAPNTTTIOFPANLQPLPGTVLIKSSGGLMLV 63
 486 VLKATGQATGAPPLVYMRASQ--AGKAPVTV--SLPASVRM--VVPQSQAGTVIGS 539
 64 SPOOT-----VTRAETSNIT--TSRP--AVPAN--PQVKTCTV 97
 540 NPKMSGAALAAAAATQKIPPSAPATMSVPACTTIVKVAATPGTTTLPATVKAASP 599
 98 -----NSSSQLIKKVAV--TPVKLAQIGTIVTVPKPSSVQSVAVPTSVTVTPKPL 150
 600 VMVSNPATRLKLAQAQVTSVSSAANTSTRPITIVKSGTV--TVAQAQVTVTVVGGTV 658
 151 NPTVTLK--PSSL--GASPTNENPL-----KENSAAVOINISPTMLEVKKCNFL-- 200
 659 KITTLVKSPTSPVSGSALISLKNKVMKVQTKPVQTSAAVGAQSTGTVQIHTKGLPA 718
 201 AMIKLACSGSQSP-----EMGQVKKLVEQLDAKLEAEFPRKLVLEKSSQPH 252
 719 GTIKLVTSADGKPTTITTTTQASGAGTKPTLIGI-----SSVSPS 759
 253 LVPLKKSVALRQLPNSQFIQOCVQOOTSMDVINTCTTVTY-----SPVVT 302
 760 TT---KGTTLTIKTIPMSALITQAGATGVTSPPGKISPTITTTKVMNSGTGAPAKIIT 816
 303 -----TVSSSQSEKSLIVSGA--TAPRT-----VSQTLNPLAGPVAK- 340
 817 AVPKRIAGHGQOQGVYVVLGAPGQPGTILRTVPMGVRVLVTEVYSAVAPATTVLVKG 876
 341 -AGVTVLHSGPTAATGATGATGLQTSKPLVTSVA--NTVTVSLOPEKPVSGTAVT 937
 877 TTGVTTLGTYTGIVST--SLAGAGAHSTASLNTPIITLCTITATLSQ---YINPAT 930
 398 LSLPVTGETSGAALCLPSKVPVSEFCMDHCKPVIGPQIKL--AOPGVLSQP 452
 931 VSAQQTLLTFAAGLTPTPTTQNPV-----SQGTQVTLTAPSGVGAQP 973

RESULT 14
 08TE50 PRELIMINARY; PRT; 1322 AA.

AC 08TE50;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
 DE EMSY protein.
 GN C10RF30.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hughes-Davies L.;
 RT "EMSY is amplified in breast cancer and displays a BRCA2 dependent DNA
 damage response."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ430203; CAD22881.1;
 SQ SEQUENCE 1322 AA; 141467 MW; 7F8C95E8BA0FC9F0 CRC64;

Query Match 8.8%; Score 197.5; DB 4; Length 1322;
 Best Local Similarity 23.0%; Pred. No. 0.00034;
 Matches 123; Conservative 78; Mismatches 197; Indels 137; Gaps 26;

5 TKVAPVAPPKVSSQGRPLPAQIVAKAPNTTTIOFPAN--LQLPG--TVLIKSSGGLML 61
 149 TTSTPTSTP-----VPGSGIATYKSPRPAS---PASVVVLPSCSTYVAKSVCSDB 197
 62 LVSPQOIVTRAFTTSNITSRPAVPANQIV---KICVFNSS--SOLIKKVA----- 108
 198 DEKPRK--RRRTNSSSSSPVLKEVPKAVPVSKITIVPVSSPKMSNIMQSIANSLLP 254
 109 -VTPVK-----KLAQIGT---VTVTVPKRSSVQSV-----AVPTSVTV 144

Db 255 HNSPVKITTFTPTOTTOTTTTOKVIYTTSSSTFVPNLSKSHNTAAVTKLVPTSVIAS 314
QY 145 TPGKPLNTVTLTKPSSL-----GASSTPSNEPN-----LKAENSA 180
Db 315 TTQKP-PVITIASQSLVSSSSSTSPPIPTAVTAVVSSTPSVMSVAGVST 373
QY 181 VOINSLPMLNKKCKNFLLMLIKLACSGSOPMGONVKKLEQLDAKIEA-EEFTT 239
Db 374 SAKMASTRLPSPKSLVSAPNOL-----AOPKHOOSPK--OQLVQOQOQOQVQAO 425
QY 240 KLYVELKSPQHLVPLFKSVVALROLPLNSOSFIOOCVOOTSDVIATCT-----TV 295
Db 426 PSPVSHQOQPOSPPLPPIKPTIQIKO-----ESGKIIITQOVOPSKILPPVATLPTS 480
QY 296 TTPSVTTTSSSSEKSIIVSGATAPRTVSOTLPLAGPVGAKAGVVT----- 345
Db 481 SNSPIVSSNCALMTKLTVTPTGQATVTRPVSPSIGMAAIPGAATVKTSSII 540
QY 346 -----LHSG-----PAAATGTTAGTGLQTSKPLVTSVANTV---TIVSLOPEKPV 391
Db 541 TVPKSLATLGGKIISNIVSGTTTKITPTMSKPNVIVQKTGKCTTIOGLPGKNV 600
QY 392 S-----CTAVTSLPVAFTFGTSGAICLPSPKPVSEFCMDHICKPV-IGTPVQ 439
Db 601 TTLNAGEKTIQIVP-----TGAKPAILTATRTITMT---VQPKGISVQ 646

RESULT 15

Q9NAS7 PRELIMINARY: PRT: 1079 AA.
AC Q9NAS7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Y51B1A.1 protein.
GN Y51B1A.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Colton M.;
RT "The sequence of C. elegans cosmid Y51B1A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006797; AAF60743.1;
DR InterPro: IPR002965; P_fich_extensn.
DR PRINTS: PRO1217; PRICHEXTENS.
SQ SEQUENCE 1079 AA; 110532 MW; 8BDE3824CF80CA1 CRC64;

Query Match 8.7%; Score 195; DR 5; Length 1079;
Best Local Similarity 22.7%; Pred. No. 0.00038;

Matches 106; Conservative 50; Mismatches 206; Indels 104; Gaps 16;

QY 5 TKVAPVSA---PPKVSQGRLLPAPQIVAVKAPNTTIOFPANLQLPCTVL-----IKSN 56
Db 399 TTTAPETTTSTPSSSTTP---VQTTITAPETTTSTPSSSTTPVQTTTITAPETTTST 454
QY 57 SGPLMLVSPQOTVTRA--ETTSNTSRPAVPANQTVKICVPSNSSLIKKVAVTVPVK 114

Db 455 EPPSSSTTPVQTTTITAPETTS--TEPPSSSTTPVQTTTITAPETTSSTPSSSTTPVQ- 511
QY 115 LAQGTIVTTVPKPSVQSAVAPTSVTVTPGKPLNVTTLKPS-----LGASS 165
Db 512 -----TTTTAPETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTITAP 565
QY 166 TSPNEPNLKAENSAVAQINLSPMLNKKCKNFLLMLIKLACSGSOPMGONVKKLVE 225
Db 566 TTSTEP---PSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTITAPETTS----- 615
QY 226 QLLDAKIAEFTTKLYVELKSSPOHLVPLFKSVVALROLPLNSOSFIOOCVOQISSD 285
Db 616 -----EPPSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTIT 654
QY 286 MVIATCT-----TIVTSPVTTTSSSSEKSIIVSGATAPRTVSOTLPLNLA 334
Db 655 APETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTP 709
QY 335 GPVGAKAGVTLHSGVP-TAATGTTAGTGLQTSKPLVTSVANTVTS 383
Db 710 -PSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTPVQTTT 768
QY 384 LQP-----EKPVSGTAV---TSLPVAFTFGTSGAICLPSPKPV 421
Db 769 TAPETTSSTPSSSTTPVQTTTITAPETTSSTPSSSTTPSSV 809

Search completed: February 16, 2003, 21:59:02
Job time : 40.7465 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 : Search time 42.3161 Seconds

(Without alignments)
1738.213 Million cell updates/sec

Title: us-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GILVTKVAPVAPPVKSSGP.....ASPTQKRIKENTSCFDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2758	100.0	801	19 AAW31494	Human hTAFII105 pr
2	2758	100.0	801	22 ABG09468	Novel human diagno
3	2758	100.0	852	21 AAY57279	Transcription fact
4	547	19.8	737	15 AAR56494	TATA-binding prote
5	547	19.8	737	17 AAW06084	Human TATA-binding
6	547	19.8	737	18 AAW25019	TATA-binding prote
7	521	18.9	1023	23 AAG82954	Human homologue of
8	256	9.3	921	15 AAR56487	TATA-binding prote
9	256	9.3	921	17 AAW06077	Drosophila TATA-bl
10	256	9.3	921	18 AAW25028	TATA-binding prote

11	256	9.3	921	22 ABB61528	Drosophila melanog
12	256	9.3	921	22 ABB66055	Drosophila melanog
13	221	8.0	2035	15 AAR57141	Host cell factor P
14	219	7.9	5179	22 AAM24516	C899P predicted am
15	214.5	7.8	1795	22 ABB69806	Drosophila melanog
16	212	7.7	842	22 ABB66631	Drosophila melanog
17	212	7.7	864	22 ABB71319	Drosophila melanog
18	211.5	7.7	1296	23 ABB66702	Drosophila melanog
19	211.5	7.7	1296	23 ABB66756	Human novel polype
20	206.5	7.5	1322	21 AAB42650	Human ORFX ORF2414
21	205.5	7.5	1920	22 AAB65656	Novel protein kina
22	205.5	7.5	2135	23 AAE21714	Human PKIN-9 prote
23	197	7.1	2972	22 AAB50363	Human SRCAP. Homo
24	197	7.1	3118	22 AAB50362	Human SRCAP. Homo
25	194.5	7.1	2781	21 AAY57453	Human transcriptio
26	194.5	7.1	2907	21 AAY57452	Human transcriptio
27	194	7.0	2971	21 AAN06725	FLO1 protein, invo
28	194	7.0	1127	22 AAB95541	Human protein sequ
29	194	7.0	1328	22 AAM78519	Human protein sequ
30	194	7.0	1331	22 AAM79503	Human protein sequ
31	194	7.0	1658	22 ABB67620	Drosophila melanog
32	194	7.0	2971	21 AAB41231	Drosophila melanog
33	193	7.0	862	15 AAR60563	Yeast 2.6 KB agglu
34	191.5	6.9	1714	22 ABB60186	Drosophila melanog
35	190.5	6.9	708	22 AAM79978	Human protein sequ
36	189	6.9	894	15 AAR47578	S. cerevisiae FLO1
37	189	6.9	894	15 AAR58754	Human protein sequ
38	188.5	6.8	709	22 AAB37063	Human protein sequ
39	188.5	6.8	709	22 AAB34334	Human protein sequ
40	188.5	6.8	709	22 AAB34334	Human protein sequ
41	188.5	6.8	881	22 AAM93811	Human polyepitide,
42	184.5	6.7	881	22 AAG70752	S cerevisiae apopt
43	184	6.6	1057	22 ABB65440	Drosophila melanog
44	183	6.6	557	22 AAB94078	Human protein sequ
45	182	6.6	2058	22 AAB97070	Human polypeptide

ALIGNMENTS

RESULT 1
AAW31494 standard; Protein: 801 AA.
AAW31494:
28-APR-1998 (first entry)
Human hTAFII105 protein.
TATA-binding protein associated binding factor 105; human; activator;
hTAFII105; transcription factor; TAF10; transcriptional activation;
antibodies; diagnosis; therapy; biopharmaceutical industry.
Homo sapiens.
US5710025-A.
20-JAN-1998.
02-OCT-1996; 96US-0725012.
02-OCT-1996; 96US-0725012.
(REGC) UNIV CALIFORNIA.
Dikstein R, Tjian R;
WPI: 1998-109818/10.
N-PSDB; AAV02872.
DNA encoding human tata-binding protein associated factor - for
producing recombinant protein

XX Claim 1; Col 17-22; 12pp; English.

CC This cDNA sequence represents a human tata-binding protein associated

CC factor, htaFII105, isolated from Daudi cell nuclear extracts. Tightly

CC associated subunits (TAF's) are components of the transcription factor

CC TFIID and are thought to mediate transcriptional activation. This encoded

CC protein may be produced recombinantly from transformed host cells or

CC purified from human cells. htaFII105 specific binding agents such as

CC specific antibodies could be used for diagnosis (e.g. genetic

CC hybridisation screens for htaFII105 transcripts), therapy (e.g. gene

CC therapy to modulate htaFII105 gene expression) and in the

CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B

CC cell specific activators or other transcriptional regulators).

XX

XX Sequence 801 AA;

SQ

Query Match 100.0%; Score 2758; DB 19; Length 801;
Best Local Similarity 100.0%; Pred. No. 2.7e-191;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTVKAVASAPKVS SSGRLPAPQIVAVKAPNTTITQIPANLQLPPTVLIKNSGRL 60
|||||
DB 1 GTLVTVKAVASAPKVS SSGRLPAPQIVAVKAPNTTITQIPANLQLPPTVLIKNSGRL 60

QY 61 MLVSPQOQTVTRAE TTNITSRPAVPANPOTVKICTVPSNSSQLIKKVAATPVKKLAQIGT 120
|||||
DB 61 MLVSPQOQTVTRAE TTNITSRPAVPANPOTVKICTVPSNSSQLIKKVAATPVKKLAQIGT 120

QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180
|||||
DB 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180

QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180
|||||
DB 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180

QY 181 VOINLSPTMLENKKCKNPLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFEFTRK 240
|||||
DB 181 VOINLSPTMLENKKCKNPLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFEFTRK 240

QY 181 VOINLSPTMLENKKCKNPLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFEFTRK 240
|||||
DB 181 VOINLSPTMLENKKCKNPLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFEFTRK 240

QY 241 LVEYELSSPOPHLVPLFKSSVALROLPLNSOSFIOQVOQTS SMDVIACTT VTTSPV 300
|||||
DB 241 LVEYELSSPOPHLVPLFKSSVALROLPLNSOSFIOQVOQTS SMDVIACTT VTTSPV 300

QY 241 LVEYELSSPOPHLVPLFKSSVALROLPLNSOSFIOQVOQTS SMDVIACTT VTTSPV 300
|||||
DB 241 LVEYELSSPOPHLVPLFKSSVALROLPLNSOSFIOQVOQTS SMDVIACTT VTTSPV 300

QY 301 VTTTSSSSOSEKSIIVSGATAPRTVSQTLNPLAGIVGAKAGVVTLSHVGPTAATGCTTA 360
|||||
DB 301 VTTTSSSSOSEKSIIVSGATAPRTVSQTLNPLAGIVGAKAGVVTLSHVGPTAATGCTTA 360

QY 301 VTTTSSSSOSEKSIIVSGATAPRTVSQTLNPLAGIVGAKAGVVTLSHVGPTAATGCTTA 360
|||||
DB 301 VTTTSSSSOSEKSIIVSGATAPRTVSQTLNPLAGIVGAKAGVVTLSHVGPTAATGCTTA 360

QY 361 GGGLQTSKPLVTSVANVTVTSILOPEKPVSGTAVTSLPAAVTFGETSGAAICLPSVRP 420
|||||
DB 361 GGGLQTSKPLVTSVANVTVTSILOPEKPVSGTAVTSLPAAVTFGETSGAAICLPSVRP 420

QY 421 VVSFCDHICKPVIQTPVOIKLAQPGPVLISOPAGITGSSKQLFSLPHVVOOPSGGNK 480
|||||
DB 421 VVSFCDHICKPVIQTPVOIKLAQPGPVLISOPAGITGSSKQLFSLPHVVOOPSGGNK 480

QY 481 OVTTTSSSTLTQKCGQKTMPTVNTIITPSFPASILKOITLPKKIISLQASPTQKNR 540
|||||
DB 481 OVTTTSSSTLTQKCGQKTMPTVNTIITPSFPASILKOITLPKKIISLQASPTQKNR 540

QY 541 IKENVTSCFRDE 552
|||||
DB 541 IKENVTSCFRDE 552

RESULT 2
ABG09468
ID ABG09468 standard; Protein: 801 AA.
XX
XX ABC09468;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #9459.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

XX Homo sapiens.

XX

XX WO200175067-A2.

XX

XX 11-OCT-2001.

PD

XX

XX 30-MAR-2001; 2001WO-US08631.

PF

XX

XX 31-MAR-2000; 2000US-0540217.

PR

XX 23-AUG-2000; 2000US-0649167.

XX

XX (HYSE-) HISEQ INC.

PA

XX

XX Drmanac RT, Liu C, Tang YT;

PI

XX

XX WPI: 2001-639362/73.

DR

XX N-PSDB; AAS73655.

DR

XX

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

PS

XX Claim 20; SEQ ID No 39827; 103pp; English.

PS

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations in

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ

Sequence 801 AA;

Query Match 100.0%; Score 2758; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 2.7e-191;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTVKAVASAPKVS SSGRLPAPQIVAVKAPNTTITQIPANLQLPPTVLIKNSGRL 60
|||||
DB 1 GTLVTVKAVASAPKVS SSGRLPAPQIVAVKAPNTTITQIPANLQLPPTVLIKNSGRL 60

QY 61 MLVSPQOQTVTRAE TTNITSRPAVPANPOTVKICTVPSNSSQLIKKVAATPVKKLAQIGT 120
|||||
DB 61 MLVSPQOQTVTRAE TTNITSRPAVPANPOTVKICTVPSNSSQLIKKVAATPVKKLAQIGT 120

QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180
|||||
DB 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180

QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180
|||||
DB 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSGASTPSNEPNLKAENSAA 180

QY 181 VOINLSPTMLENKKCKNPLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFEFTRK 240
|||||
DB 181 VOINLSPTMLENKKCKNPLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFEFTRK 240

QY 181 VOINLSPTMLENKKCKNPLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFEFTRK 240
|||||
DB 181 VOINLSPTMLENKKCKNPLAMLIKACSGSOSPENGQVKKLYEQLDLAKIEAEFEFTRK 240

QY 241 LVEYELSSPOPHLVPLFKSSVALROLPLNSOSFIOQVOQTS SMDVIACTT VTTSPV 300
|||||
DB 241 LVEYELSSPOPHLVPLFKSSVALROLPLNSOSFIOQVOQTS SMDVIACTT VTTSPV 300

Db 241 LYLKSSPOPHLVPLFKSSVVALROLPLNSQSTIQOCVOQTSSDMVATCTTTVTISPV 300
 QY 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSGPTAATGTTA 360
 CC for treating autoimmune diseases, inflammatory processes and viral or
 bacterial infections.
 XX 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSGPTAATGTTA 360
 Db 361 GTGLQTSKPLVTSVANTVTTSVLSQPEKPVVSGTAVTSLPAVFGETSGAATCLPSVKP 420
 QY 361 GTGLQTSKPLVTSVANTVTTSVLSQPEKPVVSGTAVTSLPAVFGETSGAATCLPSVKP 420
 Db 361 GTGLQTSKPLVTSVANTVTTSVLSQPEKPVVSGTAVTSLPAVFGETSGAATCLPSVKP 420
 QY 421 VVSCWMDHICKPVIGTPOIKLAOPGVLSPAGIPGSSSKOLFSLFHVVOQPSGGNEK 480
 Db 421 VVSCWMDHICKPVIGTPOIKLAOPGVLSPAGIPGSSSKOLFSLFHVVOQPSGGNEK 480
 QY 481 QVTTISHSSTLTIOKCGKTPVNTIIPTSQPPASILKQITLPGNKILSLQASPTQKNR 540
 Db 481 QVTTISHSSTLTIOKCGKTPVNTIIPTSQPPASILKQITLPGNKILSLQASPTQKNR 540
 QY 541 IKENVTSQCFRDE 552
 Db 541 IKENVTSQCFRDE 552

RESULT 3

AAR57279
 ID AAY57279 standard; Protein: 852 AA.

AAY57279;

06-JUN-2000 (first entry)

DE Transcription factor subunit TAFII105 polypeptide.

KM TATA box-binding protein associated factor II 105; TAFII105; cancer;
 KM transcription factor; apoptosis; cytoskeletal; immunosuppressive;
 KM antiinflammatory; viricide; antibacterial.

OS Homo sapiens.

XX Key Location/Qualifiers

FH Misc-difference 802

FT /label= Glx

FT /note= "encoded by TGA"

FT Misc-difference 834

FT /label= Glx

FT /note= "encoded by TAA"

XX W0200012699-A1.

PN 09-MAR-2000.

PD 25-AUG-1999; 99W0-IL00464.

PF 27-AUG-1998; 98IL-0125971.

PR (YEDA) YEDA RES & DEV CO LTD.

PA Dikstein R, Yamit-hezi A;

PI WPI: 2000-256640/22.

XX N-PSDB; AAZ90465.

DR Polypeptide encoding TATA box binding protein associated factor II 105

XX useful for treating e.g. cancers and inducing apoptosis has a dominant

PT negative effect on the normal biological activity of the binding

PT protein -

XX Claim 7; Fig 2; 48pp; English.

XX This represents a polypeptide comprising a (modified) fragment (I) of

CC a TATA box-binding protein associated factor II 105 (TAFII105). A

CC pharmaceutical composition comprising (I) or the polynucleotide or an

CC inhibitor or antagonist of (I) is useful for treating cancers and

CC inducing apoptosis in pathological cells. The composition is also useful

CC for treating autoimmune diseases, inflammatory processes and viral or

CC bacterial infections.

XX Sequence 852 AA;

Query Match 100.0%; Score 2758; DB 21; Length 852;

Best Local Similarity 100.0%; Pred. No. 3e-191;

Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTLVTKAVPAPVAPKRVSSGPRLPAPQIYAVKAPNTTIQFPANLQLPFGYVLIKNSGPL 60
 Db 1 GTLVTKAVPAPVAPKRVSSGPRLPAPQIYAVKAPNTTIQFPANLQLPFGYVLIKNSGPL 60
 QY 61 MLVSPQQTVAETNTSNTSRPAPANPQYKICVPPSSSOLIKKAVTVPKKLAQIGT 120
 Db 61 MLVSPQQTVAETNTSNTSRPAPANPQYKICVPPSSSOLIKKAVTVPKKLAQIGT 120
 QY 121 TVVTVPRKSSVQSVAVPTSVVTPGKPLNTVTTLKPSISGASTSPNEPNLKAENSA 180
 Db 121 TVVTVPRKSSVQSVAVPTSVVTPGKPLNTVTTLKPSISGASTSPNEPNLKAENSA 180
 QY 181 VQINLSPMLENVKCKKFLMLIKLACSGSOSPEMGQNVKKLYQOLIDAKIEAEFTRK 240
 Db 181 VQINLSPMLENVKCKKFLMLIKLACSGSOSPEMGQNVKKLYQOLIDAKIEAEFTRK 240
 QY 241 LVELKSSPOPHLVPLFKSSVVALROLPLNSQSTIQOCVOQTSSDMVATCTTTVTISPV 300
 Db 241 LVELKSSPOPHLVPLFKSSVVALROLPLNSQSTIQOCVOQTSSDMVATCTTTVTISPV 300
 QY 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSGPTAATGTTA 360
 Db 301 VTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVYTLHSGPTAATGTTA 360
 QY 361 GTGLQTSKPLVTSVANTVTTSVLSQPEKPVVSGTAVTSLPAVFGETSGAATCLPSVKP 420
 Db 361 GTGLQTSKPLVTSVANTVTTSVLSQPEKPVVSGTAVTSLPAVFGETSGAATCLPSVKP 420
 QY 421 VVSCWMDHICKPVIGTPOIKLAOPGVLSPAGIPGSSSKOLFSLFHVVOQPSGGNEK 480
 Db 421 VVSCWMDHICKPVIGTPOIKLAOPGVLSPAGIPGSSSKOLFSLFHVVOQPSGGNEK 480
 QY 481 QVTTISHSSTLTIOKCGKTPVNTIIPTSQPPASILKQITLPGNKILSLQASPTQKNR 540
 Db 481 QVTTISHSSTLTIOKCGKTPVNTIIPTSQPPASILKQITLPGNKILSLQASPTQKNR 540
 QY 541 IKENVTSQCFRDE 552
 Db 541 IKENVTSQCFRDE 552

RESULT 4

AAR56494
 ID AAR56494 standard; Protein: 737 AA.

AAR56494;

23-MAR-1995 (first entry)

DE TATA-binding protein-associated factor hTAFII130.

KM TATA-binding protein associated factor; hTAFII130; screening;

KM diagnostic; therapeutic; gene transcription regulation.

OS Homo sapiens.

XX W09417087-A.

PN 04-AUG-1994.

PD 28-JAN-1994; 94W0-US01114.

PF 28-JAN-1993; 93US-0013412.

XX 28-JAN-1993; 93US-0013412.


```

XX PF 28-JUN-2001; 2001MO-US20592.
XX XX
XX PR 29-JUN-2000; 2000US-215164P.
XX PR 10-AUG-2000; 2000US-224457P.
XX PA (ANAD-) ANADYS PHARM INC.
XX XX
XX PI Moore J, Buurman ET, Desliva T, Harris S, Komaritsky S;
XX PI Mendillo M, Moore D, Mccoy M, Sanderson K, Haq T, Zhu S, Long F;
XX PI Davidov E, Thompson CM;
XX DR MPI: 2002-147962/19.
XX DR N-PSDB; ABR32842.
XX XX
XX PT Screening candidate antifungal compound for interaction with essential
XX PT protein, modulation of essential protein activity, binding to essential
XX PT protein, by contacting protein with test compound and determining
XX PT effects -
XX PS Claim 1; Figure 79; 522pp; English.
XX XX
XX CC The invention describes a method of screening a candidate antifungal
XX CC compound for interaction with essential proteins (EP) or for modulation
XX CC of EP activity e.g. fungal gene transcription. The proteins tested in the
XX CC invention include RBC34, POB3, TRN2, NAB2, MPT1, MTR2, BOS1, POB30, RSN2,
XX CC SQ11, MRM1, TRB1, SPC98, BFR2, RNAL, GCD7, SKI6, NRP1, LCP5, NCE103,
XX CC ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans
XX CC and human homologues. The method involves contacting a culture with one
XX CC or more test compounds and determining the effects on the growth or
XX CC viability of cells of the culture which preferably comprises fungal cells
XX CC or yeast cells. Preferably the identified compounds interact with, or
XX CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor
XX CC compounds identified by the method are useful for preventing or
XX CC inhibiting fungal, particularly C. albicans growth in culture or in a
XX CC mammal. The antifungal agents interact with essential fungal elements
XX CC that can be used to treat fungal infection by preventing the growth and
XX CC preferentially killing the fungi, but does not inhibit the biological
XX CC activity of mammalian homologues. This amino acid sequence represents a
XX CC target protein used to test the antifungal compounds, described in the
XX CC method of the invention.
XX SQ Sequence 1023 AA;

Query Match 18.9%; Score 521; DB 23; Length 1023;
Best Local Similarity 31.7%; Pred. No. 5,5e-29;
Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;

OY 13 PKVSSG-----RLPAPQIVAAKAPPTTTIOPANLQPPGTVLIRKNSGPILM 61
DB 414 PRATTSIGIRATLPTVLAIRLP-----PQNPNTNIO--NFQLPPGMVLVNSENGQL 464
OY 62 LVSPQQTVR-----AETTSNTTSRAVPANPQTVKICTPVNSSQLIKKVAATEVKKL 115
DB 465 MI-POQALQMOQAHAQPOQTWAPRATPTSNAPVOISTVQAPGPIIAR-QVTP----- 518
OY 116 AQIGTVTVTVPKPSSVGSVAVPTSVVTVPKPLMTV--TLKPSLSGASS-----TFS 168
DB 519 ----TTIINQV--SQAQTTVQPSATLQRSFGVQPOLVGAQAQTSLCTATRAVQGTPO 571
OY 169 NE-PNLKAENSAVAOINLSPMLNKKKCNFLAMLIKLAGSSQPEKQONKRLVEOD 227
DB 572 KRVPGATTTSSAATE-----TMENYKCKCNFLSTIKLAGSSQOSTETRAANKVELYONL 625
OY 228 LDKATIAEEFTKRLVELKSSQPHLVPLKRSVALROLPLNPSQSFIOOCVQO-----TS 283
DB 626 LDKGIAEAEFTSRLYRELNSSQPHLVPLKRSIPALROLTPDSAFIOOSQOQPPPPS 685
OY 284 SDMVATCTTYYTSPVVTTVSSSQSEKSIIVSGATARTVSVOTLNPLAGVYGAACV 343
DB 686 Q-----ATTLTAVALVSSSVQRTAGKTAATVTSALQPPVLSL----- 722
OY 344 VTLHSVPTAAAGTGTAGTGLQTSKPLVTSVANTVTVSLQEPKPVSGTAVTSLPAAV 403

```

```

DB 723 -----TQPTQVGKQKQGPPTPLVYQ-----QPKP-----GALINPOV 756
OY 404 TFGETSGMAICLPVSVKPVVSCMDHICKPVITGPVOIKLAQPGPVL-----SQPAGIPTGS 459
DB 757 TLVQT-----PMVALRQPH-NRIMLTTPQGVNISEEARIATNSELVGLTNS 804
OY 460 SSQQLSLEFHVVO-----QPSGNEKO---VTTISHSSTLTIOKCGOK 499
DB 805 CKDETFLQAPLQRIILEIGKKHGITELHPDVSVSVSHATQRIQLNLYEK 854

RESULT 8
AARS6487
ID AARS6487 standard; Protein; 921 AA.
XX AARS6487;
AC AARS6487;
XX 23-MAR-1995 (first entry)
XX DT
XX DE TATA-binding protein-associated factor dTAFl110.
XX KW TATA-binding protein associated factor; dTAFl110; screening;
XX KM diagnostic; therapeutic; gene transcription regulation.
XX OS Drosophila.
XX PN W09417087-A.
XX PD 04-AUG-1994.
XX PF 28-JAN-1994; 94MO-US01114.
XX PR 28-JAN-1993; 93US-0013412.
XX PR 30-JUN-1993; 93US-0087119.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
XX PI Tjian R, Wang E, Weinzierl ROJ;
XX DR MPI: 1994-264019/32.
XX DR N-PSDB; AAO70724.
XX PT TATA-binding protein associated protein factors - and
XX PT corresponding nucleotide sequence and deriv. antibodies, useful
XX PT in screening, diagnostics and therapeutics
XX PS Disclosure; Page 56-61; 180pp; English.
XX CC The TATA-binding protein associated factor dTAFl110 (including
XX CC specific antibodies and fusion products) are used in drug screening,
XX CC diagnostics and therapeutics. They are used in the development of
XX CC specific biochemical assays for screening compounds that agonise or
XX CC antagonise selected transcription factors involved in regulating
XX CC gene expression associated with human pathology.
XX SQ Sequence 921 AA;

Query Match 9.3%; Score 256; DB 15; Length 921;
Best Local Similarity 22.9%; Pred. No. 7.6e-10;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

OY 25 PQIVAAKAPNTTTIOPANLQPPGTVLIRKNSGPILM-----VSPQQTVTRAETTS 76
DB 133 PQSPSTLSTLNTNGQRA-----LVKTDNGFQLLVGTTGEPYVTOITNTNSNS 184
OY 77 NITSRAVPANPQTVKICTPVNSSQ-----LIKRVAVTPVKKLAQIGTVVTVTP 127
DB 185 NITSTNNHPTTQ-IRLQTVPAASMTNTTANSIIVNSVASSGVANSQPHLQNLNQ 243
OY 128 KPSSVSAVAVPTSVVTVTGKPLNTVTLKPSLSGASSPSPNEPNKAENSAVAOINLSP 187

```

```

Db 244 APOLPOTIOTIIPAOSQOOOVNNSAGCTATAVSSSTA-----ATT 287
QY 188 TMLENVK-KCKNPLAMLIKACSGSPENGONVKILEOLLDAKIEAEFTRLKLYELK 246
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 288 TGGNTREKCKRFLANLIEL--STREPKPEKKNRTLIQELVANNVPEECDELERLN 345
QY 247 SSPOPHLVPLKSSVALROL-----LPSQSFTQ--- 276
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 346 ASPQCLIGFLKSLPRLROALYKELVIEGIRPPQHVLAGLSQLPKIQAOIRPIG 405
QY 277 ----QCVOQTSSDMVIATCTTTVTTSPTVTTVSSSSEKSIIVSGATAPRTVS---VQT 329
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 406 PSQTTTIGQOVNRM--TPNALGTPRPITGHTTISKQPPN---IRLPAPRLVMTGIRT 460
QY 330 LNPAGVGAAGAGVYTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVTVVLOPEKP 389
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 461 QIP-SLOVPGQANIVQIR--GPOHLOLORTSGVOIRATTRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTLSLPAVTFEGTSGAATCLPSVKPVVSCMDHICKPVIGTPVOIKLAQPGVL 449
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 507 -----KLTAVKVQOTOIKAI-TPSLHP-----PSL 530
QY 450 SOPAGIPTGSSSKQLFSLFHVYQOPSGGNEKQVTTISHSSTLTQKCGOKTMAPVTIIP 509
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 531 AAISGGPPPTPLSVLS-----TLNSAST-----TTLPIPS-LPT 564
QY 510 SOPPPASI---LKQITLPGN-----KILSLQA---SPTOKNRKENV 546
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 565 VHLPEALRLAREQOMQNSLNHNSNHFDAKLVEIKAPSLHPIMERINMSLT 614

```

RESULT 9
AAM06077 standard; Protein: 921 AA.

XX AAM06077;

DT 27-JAN-1997 (first entry)

DE Drosophila TATA-binding protein associated factor dTAFII110 protein.

KW Drosophila; TATA-binding protein; TBP associated factor; TFIID;

KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;

XX holoenzyme; lambda-gt11; expression library.

XX Drosophila melanogaster.

XX US5534410-A.

XX 09-JUL-1996.

XX 28-JAN-1993; 93US-0013412.

XX 28-JAN-1994; 94US-0188582.

XX 30-JUN-1993; 93US-0013412.

XX (REGC) UNIV CALIFORNIA.

XX Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;

XX Tjian R, Wang E, Weinzierl ROJ;

XX MPI; 1996-333245/33.

XX N-PSDB; AAT42210.

XX Screen for cpds. that bind human TATA-binding protein associated

XX factor - by testing ability to bind to polypeptide fragments of the

XX factor, useful as (ant)agonists of transcription factors involved in

XX disease.

CC (TBP) associated factor (TAF) designated TAFII110. The protein is a
CC component of the TFIID fraction required for reconstituting RNA
CC polymerase II in vitro transcription activity. The encoded protein
CC has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.
CC based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.
CC The invention relates to purified proteins involved in transcription
CC by RNA polymerase II, the RNA polymerase which transcribes messenger
CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
CC contains a TBP and other TAFs. Purification of TFIID and separation of
CC its components reveals 7 proteins ranging in size from 30-250 kD.
CC Serum raised against the TFIID fraction allowed cloning of the corresp.
CC genes from lambda-gt11 expression libraries.

CC Sequence 921 AA;

CC Query Match 9.38; Score 256; DB 17; Length 921;

CC Best Local Similarity 22.9%; Pred. No. 7.6e-10;

CC Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

```

QY 25 PQIYAVKAPNTTTOEFPANLQLPQGVTLIKNSGPIML-----VSPOQTVRAEFTS 76
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 133 PPSPIITLSTNTGOTPA-----LIVKTDNGQLRVCTTGPPVOTITFTSNNS 184
QY 77 NITSRAVPANPQVTKICTVFNSSQ-----LIKKVAVTPYKLAQIGTVVTVTP 127
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 NTSSTNHPPTTQ--IRLQTPAASMTNTTATSNITVNSVSSGANSOPPHLQUNAQ 243
QY 128 KPSSVQSAVPTSVYVTPGKPLNTVTTLKPSLSLCASTPSNEPRLKENSANAQINISLP 187
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 244 APOLPOTIOTIIPAOSQOOOVNNSAGCTATAVSSSTA-----ATT 287
QY 188 TMLENVK-KCKNPLAMLIKACSGSPENGONVKILEOLLDAKIEAEFTRLKLYELK 246
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 288 TGGNTREKCKRFLANLIEL--STREPKPEKKNRTLIQELVANNVPEECDELERLN 345
QY 247 SSPOPHLVPLKSSVALROL-----LPSQSFTQ--- 276
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 346 ASPQCLIGFLKSLPRLROALYKELVIEGIRPPQHVLAGLSQLPKIQAOIRPIG 405
QY 277 ----QCVOQTSSDMVIATCTTTVTTSPTVTTVSSSSEKSIIVSGATAPRTVS---VQT 329
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 406 PSQTTTIGQOVNRM--TPNALGTPRPITGHTTISKQPPN---IRLPAPRLVMTGIRT 460
QY 330 LNPAGVGAAGAGVYTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVTVVSLQPEKP 389
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 461 QIP-SLOVPGQANIVQIR--GPOHLOLORTSGVOIRATTRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTLSLPAVTFEGTSGAATCLPSVKPVVSCMDHICKPVIGTPVOIKLAQPGVL 449
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 507 -----KLTAVKVQOTOIKAI-TPSLHP-----PSL 530
QY 450 SOPAGIPTGSSSKQLFSLFHVYQOPSGGNEKQVTTISHSSTLTQKCGOKTMAPVTIIP 509
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 531 AAISGGPPPTPLSVLS-----TLNSAST-----TTLPIPS-LPT 564
QY 510 SOPPPASI---LKQITLPGN-----KILSLQA---SPTOKNRKENV 546
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 565 VHLPEALRLAREQOMQNSLNHNSNHFDAKLVEIKAPSLHPIMERINMSLT 614

```

RESULT 10
AAM25028 standard; Protein: 921 AA.

XX AAM25028;

XX 08-OCT-1997 (first entry)

XX TATA-binding protein associated factor, dTAFII110.

XX TATA-binding protein associated factor; TAF; nuclear protein;

XX

Db 133 PQSPITLSTLNTGOTPA-----LLVKTNGFQLLRVGTTGPTVTOTITNTSNN 184
 QY 77 NITSRPAVPANPQTVKICTVPSNSSQ-----LIKKAVTPYKLAQIGTVVTVTP 127
 Db 185 NITSTTNHPTTQ-IRLQTVPAASMTNTATSNIVNSVSSGANSQPHPLQLNAQ 243
 QY 128 KPSSVQSAVAVPTSVVTPGKPLNTVTLKPSLSGASSTPSNEPLKAKNSAAVOINIS 187
 Db 244 APQLPQIOTIOTIPAQSQOQOVNNSVAGATAVSSSTA-----ATT 287
 QY 188 TMLBNVK-KCKNFLAMLKILKACSSQSPKMGONKVLBOLLDAKIEAEFTKRLYVEIK 246
 Db 288 TQGNTEKCKKFLANLIEL-STREKPVKKNVTLIQELVNAVNEPEECDRLERLIN 345
 QY 247 SSPQHLVPLKKSVALROL-----LPSNSQSFQ--- 276
 Db 346 ASPQCLIGFLKSLPLRLQALYKELVIEGKPPQHVLAGLSQQLPKIQAOIRPIG 405
 QY 277 ----QCVQOTSSDMVIACTTTVTTSPPVYTTVSSQSEKSIYSGATAPRTVS---VQT 329
 Db 406 -PSQTTTIGQTOVRMI--TPNALGTPRPRTIGHTTISKOPPN--IRLPTAPRLVNTGIRT 460
 QY 330 LNPAGPVGAKAGVYTLHVSPTATGTTAGTGLTOSKPLVTSVANTVTVSIOPEKP 389
 Db 461 QIP-SLOVPGQANIVQIR--GPOHAQLORTGSVOIRATRP-----PNSVPTAN----- 506
 QY 390 VVSGTAVTLSLPAVTFGTSGAALCPVSKPVVSFCMDHICKPVIGTPVQIKLAOPGVL 449
 Db 507 -----KLTAVKVGOTOIKAI-TPSLHP-----PSL 530
 QY 450 SOPAGIPTGSSSKOLFSLFHVVOOPSGENKQVTTISHSSTLTIOKCGOKTAPVNTIIP 509
 Db 531 AATSGPPTPTLSVLS-----TLNAST-----TLPIPS-LPT 564
 QY 510 SQFPFASIT---LKOITLPGN-----KILSLQA---SPTQKNRIKENVT 546
 Db 565 VHLPEALRAREOMONSLNHSNHFDAKLVETKAPSLHPHMERINAST 614
 RESULT 12
 ID ABB66055 standard; Protein: 921 AA.
 AC ABB66055:
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 24957.
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX Drosophila melanogaster.
 OS
 XX
 PN W0200171042-A2.
 PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL10158.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure: SEQ ID NO 24957; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AAR5737-AAR72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 921 AA;
 SQ
 Query Match 9.3%; Score 256; DB 22; Length 921;
 Best Local Similarity 22.9%; Pred. No. 7.6e-10;
 Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;
 QY 25 PQIVAKAPNPTTIOFPANLQLPCTVLKISNGPLML-----VBPQOTVTAETTS 76
 Db 133 PQSPITLSTLNTGOTPA-----LLVKTNGFQLLRVGTTGPTVTOTITNTSNN 184
 QY 77 NITSRPAVPANPQTVKICTVPSNSSQ-----LIKKAVTPYKLAQIGTVVTVTP 127
 Db 185 NITSTTNHPTTQ-IRLQTVPAASMTNTATSNIVNSVSSGANSQPHPLQLNAQ 243
 QY 128 KPSSVQSAVAVPTSVVTPGKPLNTVTLKPSLSGASSTPSNEPLKAKNSAAVOINIS 187
 Db 244 APQLPQIOTIOTIPAQSQOQOVNNSVAGATAVSSSTA-----ATT 287
 QY 188 TMLBNVK-KCKNFLAMLKILKACSSQSPKMGONKVLBOLLDAKIEAEFTKRLYVEIK 246
 Db 288 TQGNTEKCKKFLANLIEL-STREKPVKKNVTLIQELVNAVNEPEECDRLERLIN 345
 QY 247 SSPQHLVPLKKSVALROL-----LPSNSQSFQ--- 276
 Db 346 ASPQCLIGFLKSLPLRLQALYKELVIEGKPPQHVLAGLSQQLPKIQAOIRPIG 405
 QY 277 ----QCVQOTSSDMVIACTTTVTTSPPVYTTVSSQSEKSIYSGATAPRTVS---VQT 329
 Db 406 -PSQTTTIGQTOVRMI--TPNALGTPRPRTIGHTTISKOPPN--IRLPTAPRLVNTGIRT 460
 QY 330 LNPAGPVGAKAGVYTLHVSPTATGTTAGTGLTOSKPLVTSVANTVTVSIOPEKP 389
 Db 461 QIP-SLOVPGQANIVQIR--GPOHAQLORTGSVOIRATRP-----PNSVPTAN----- 506
 QY 390 VVSGTAVTLSLPAVTFGTSGAALCPVSKPVVSFCMDHICKPVIGTPVQIKLAOPGVL 449
 Db 507 -----KLTAVKVGOTOIKAI-TPSLHP-----PSL 530
 QY 450 SOPAGIPTGSSSKOLFSLFHVVOOPSGENKQVTTISHSSTLTIOKCGOKTAPVNTIIP 509
 Db 531 AATSGPPTPTLSVLS-----TLNAST-----TLPIPS-LPT 564
 QY 510 SQFPFASIT---LKOITLPGN-----KILSLQA---SPTQKNRIKENVT 546
 Db 565 VHLPEALRAREOMONSLNHSNHFDAKLVETKAPSLHPHMERINAST 614
 RESULT 13
 ID AAR57141 standard; Protein: 2035 AA.
 AC AAR57141:
 XX 19-MAR-1995 (first entry)
 DE Host cell factor protein.
 KW Herpes simplex virus; herpes virus; VP16; immediate early gene;
 XX

CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB85737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1795 AA;

Query Match 7.8%; Score 214.5; DB 22; Length 1795;
Best Local Similarity 20.2%; Pred. No. 2e-06;
Matches 130; Conservative 86; Mismatches 254; Indels 173; Gaps 22;

QY 9 PVSAPPKVSSGPRLPAPQIVAAVAPNTTIIQFPANLQ-----PPGVLIKNSNG 58
DB 700 PMSSTCKPTTTRK-PSTRITPTTKYTTTQITTTPLRSSSTETTSQPPPTTTPQPTT 758
QY 59 PLMLVSPQOTVTRAEFTSNITSRPAVPANPQYKICTVPNSSQLIKYAVTPVKRLAQI 118
DB 759 TLTLYTPKSTTTTTEKPTSSPK-PTTQKWTSTAPNWT-----KVAITQKETTP 812
QY 119 GTT-----VTTVPKRS-----VQSAVPISVTVTPGKPLNT 152
DB 813 QSTSTTIFRKTTNNPEPTSTEKPTSTPKSTTPPKTSTVASTSEKTISSPKPTE 872
QY 153 VTTLKPSLSGASTPNEPNLKAENSAVAQINSPTMLENVKCKNFMLMLIKLACSGSQ 212
DB 873 KSTENFTTNSVTSALTSSTQRA-----TSTSEPTKTONITTTTPKPTLKTS 922
QY 213 SPEMGONVKLVEQLLDAN--IEAEFTRLKLYELKSSPOPHLVPLKRSVALROLPN 270
DB 923 TOEATTSQKVSATVITTKKATFESSPLTLSTEEPNTPKP-----LRTTTP 970
QY 271 SOSFIOQCVOQSSDMVATCTTVTTSVVTTSVSSQSEKIIIVSGA--TAPR--TVS 326
DB 971 TTS-----VTATRTITTTTISESTETSTOKPKSTPTSTRTTPKVTYVA 1017
QY 327 VOTLNLAPGVGAKAGVTLHSVG-----PTAATGTTA--GTGLLOTSKPL 371
DB 1018 VSTQNPNT--TTSKSTVTITTPNPSPSTQRPPTTTRQPTSTASTSICITRIPTTNP 1075
QY 372 VTSVANTVTTSLOPEK-----VSGTAVTLSLPAVTFGETSGAICLPSPKPVASF 424
DB 1076 QNSTSTDLTYTRPPCPDSDTSDKNTWTACTQELQVNLLE-----LQSPKQEOF 1128
QY 425 CWDHICKPIGT-----PVQIKLAQPGVLSQPAIGTSSSKOLFSLF 468
DB 1129 THTRHTALTGSRNTLGGQEVDPDYMDAPSSAEESGQATTAAPTMSTLAANAHLQKLF 1188
QY 469 HVV-----QPSGNEKQVTT-----ISHSFLTIQKCG 497
DB 1189 HIISTTPSRERHAPTORPSQSSQSRSGVTLAOMARHNLATSKPFIASHLSLQOLA 1248
QY 498 ---OKTMPVNTII-----PTSQFPASILKQITLPGNKII 529
DB 1249 STOKRSIIPKTLVTNHTTKPEDESEYDSETSQYTDENDEVL 1291

Search completed: February 16, 2003, 21:54:51
Job time : 55.3161 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 14.1842 seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GTLVTKVAPVSAAPRVSSGP.....ASPTOKRIKENVSCFRDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/ECTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backtilist1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2758	100.0	801	1	US-08-725-012-2
2	547	19.8	737	1	US-08-188-582-16
3	547	19.8	737	1	US-08-188-582-16
4	256	9.3	921	1	US-08-188-582-2
5	256	9.3	921	1	US-08-188-582-2
6	221	8.0	2035	1	US-08-046-585-5
7	221	8.0	2035	1	US-08-046-585-5
8	221	8.0	2035	1	US-08-393-703-5
9	221	8.0	2035	5	PCT-US93-11721-5
10	197	7.1	2972	4	US-09-579-181-2
11	193	7.0	862	1	US-08-325-267A-4
12	189	6.9	894	3	US-08-362-525-22
13	189	6.9	894	3	US-08-362-525-22
14	179.5	6.5	1537	1	US-08-325-267A-2
15	174.5	6.3	1721	3	US-08-700-651-5
16	174.5	6.3	1721	3	US-08-700-651-5
17	172	6.2	750	4	US-09-165-239A-4
18	166	6.0	805	4	US-09-103-429A-4
19	164.5	6.0	752	1	US-08-244-189-2
20	164.5	6.0	1837	3	US-08-828-361B-5
21	164	5.9	903	2	US-08-853-310-2
22	162	5.9	3969	4	US-08-061-376-5
23	158.5	5.7	786	4	US-09-103-429A-3
24	157.5	5.7	941	4	US-07-757-022B-14
25	157.5	5.7	1022	4	US-07-757-022B-84
26	157.5	5.7	1038	4	US-07-757-022B-74
27	157.5	5.7	1049	4	US-07-757-022B-58

28	157.5	5.7	1140	4	US-07-757-022B-104	Sequence 104, App
29	157.5	5.7	1270	4	US-07-757-022B-44	Sequence 44, App1
30	157.5	5.7	1311	4	US-07-757-022B-42	Sequence 42, App1
31	157.5	5.7	1313	4	US-07-757-022B-142	Sequence 142, App
32	157.5	5.7	1314	4	US-07-757-022B-50	Sequence 50, App1
33	157.5	5.7	1320	4	US-07-757-022B-46	Sequence 46, App1
34	157.5	5.7	1354	4	US-07-757-022B-60	Sequence 60, App1
35	157.5	5.7	1361	4	US-07-757-022B-48	Sequence 48, App1
36	157.5	5.7	1363	4	US-07-757-022B-40	Sequence 40, App1
37	157.5	5.7	1363	4	US-07-757-022B-52	Sequence 52, App1
38	157.5	5.7	1404	4	US-07-757-022B-2	Sequence 2, App11
39	157.5	5.7	1404	4	US-07-757-022B-62	Sequence 62, App11
40	155	5.6	907	3	US-08-783-774-2	Sequence 2, App11
41	155	5.6	907	4	US-09-328-599A-1	Sequence 1, App11
42	155	5.6	907	5	PCT-US95-04611A-19	Sequence 19, App1
43	153	5.5	878	4	US-09-556-706B-2	Sequence 2, App11
44	151.5	5.5	1125	4	US-09-513-783A-152	Sequence 152, App
45	151.5	5.5	1610	4	US-09-513-783A-22	Sequence 22, App1

ALIGNMENTS

```
RESULT 1
US-08-725-012-2
; Sequence 2, Application US/08725012
; Patent No. 5710025
; GENERAL INFORMATION:
; APPLICANT: Dikstein, Rivka
; APPLICANT: Tjian, Robert
; TITLE OF INVENTION: B-Cell Specific Transcription Factor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,012
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-725-012-2

Query Match      100.0%; Score 2758; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 1.9e-223;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GTLVTKVAPVSAAPRVSSGPLPAPQIVAAKAPVTTTIOPPANIQLPFGVTLKNSGGL 60
      |||||||
Db      1 GTLVTKVAPVSAAPRVSSGPLPAPQIVAAKAPVTTTIOPPANIQLPFGVTLKNSGGL 60
      |||||||
OY      61 MLVSPQQTVAETTSNITSRPVAPVAPNAPQVVKICTVBNSSQLTKVAVTPVKKLAQIGT 120
      |||||||
```

```

Db 61 MLVSPQGVTRAEITNSITSRPVPANPOTVAKICTVPSNSSOLIKKAVTPVKKIAQIGT 120
QY 121 TVTVPKPSVQSVAVPTSVTVTPGKPLNTVTTLKPSLGCASSTPSNEPLKKAENSA 180
Db 121 TVTVPKPSVQSVAVPTSVTVTPGKPLNTVTTLKPSLGCASSTPSNEPLKKAENSA 180
QY 181 VOINISPTMLEVKKCKNFKLMLIKLACSGSSPEMGQVKKLVLEQLDAKIEAEFTTK 240
Db 181 VOINISPTMLEVKKCKNFKLMLIKLACSGSSPEMGQVKKLVLEQLDAKIEAEFTTK 240
QY 241 LVELKSSPQPLVPLFKKSVALLROLPLNSQSFIOQCVOQTSSPDVIACTTTVTTSV 300
Db 241 LVELKSSPQPLVPLFKKSVALLROLPLNSQSFIOQCVOQTSSPDVIACTTTVTTSV 300
QY 301 VTTVSSSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGVTLHSVPTAATGCTTA 360
Db 301 VTTVSSSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGVTLHSVPTAATGCTTA 360
QY 361 GTGLQTSKPLVTSANVTVTYSLOPEKRVSGTAVTSLPVTGETSGAICLPSTVP 420
Db 361 GTGLQTSKPLVTSANVTVTYSLOPEKRVSGTAVTSLPVTGETSGAICLPSTVP 420
QY 421 VVSEFCMDHICKRVIGTPOVIAKLOPQVLSQAGIPTGSSKQLFSLFHVQOQSGGNEK 480
Db 421 VVSEFCMDHICKRVIGTPOVIAKLOPQVLSQAGIPTGSSKQLFSLFHVQOQSGGNEK 480
QY 481 QVTTTSHSSTLTQKCGQKTMVNTIIPTSQPPASILKQITLPGNKILSLQASPTQKR 540
Db 481 QVTTTSHSSTLTQKCGQKTMVNTIIPTSQPPASILKQITLPGNKILSLQASPTQKR 540
QY 541 IKENTSCFRDE 552
Db 541 IKENTSCFRDE 552

```

RESULT 2
 US-08-188-582-16
 Sequence 16 Application US/08188582
 Patent No. 5534410
 GENERAL INFORMATION:
 APPLICANT: Tjian, Robert
 APPLICANT: Comal, Lucio
 APPLICANT: Dynlacht, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoko
 APPLICANT: Wang, Edith
 APPLICANT: Weinzierl, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS.
 TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TARS AND METHODS OF USE
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/188,582
 FILING DATE: 28-JAN-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A.
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989

```

;
; TELEFAX: (415) 398-3249
;
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-188-582-16

```

Query Match 19.8%; Score 547; DB 1; Length 737;
 Best Local Similarity 30.2%; Pred. No. 1.4e-37;
 Matches 173; Conservative 67; Mismatches 144; Indels 186; Gaps 21;

```

QY 13 PKRVSSG-----PRLPAQIVAAKAPNTTITQEPANIQLPPTGLKSNGLM 61
Db 68 PRATTSGRINATLPTVPLARLP-----PQPMNIO--NQLPGRVVLVSENGQL 118
QY 62 LVSPOQTVTR-----AETTSNITSRPANVPANPQVTKICTVPSNSSOLIKKAVTPVKKL 115
Db 119 MI-PQALAQMOQAHAQOPQTMAPAPPTSPAPVQISTVQAPGPIIAR-QVTP----- 172
QY 116 AQIGTVVTVTPKPSVQSVAVPTSVTVTPGKPLNTVT--TLKPSLGCAS-----TPS 168
Db 173 -----TTIIOV---SQAQTTVPQSATLQSPGVQPOLVGGAAQTASLGTATAVOTGTPO 225
QY 169 NE-PNLKAENSAVAQINLSPTMLENVKKCKNFKLMLIKLACSGSSPEMGQVKKLVLEQL 227
Db 226 RVVPGATTTSSAATE-----TWENVKCKCKNFKLMLIKLACSGKOSTETAAVKKELVQL 279
QY 228 LAKIEAEFTKRLVLELKSSPQPLVPLFKKSVALLROLPLNSQSFIOQCVOQ-----TS 283
Db 280 LDKIEAEFTSKRLVLELKSSPQPLVPLFKKSVALLROLPLNSQSFIOQCVOQPPPPPS 339
QY 284 SDKVIATCTTVTTSPTVTVSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
Db 340 Q-----ATFALTAVLVSSSVQRTAGKTAAYTSLQPPVLS----- 376
QY 344 VTLHSVPTAATGCTTAGTGLQTSKPLVTSANVTVTYSLOPEKRVSGTAVTSLP 403
Db 377 -----TOPQVGVGKQGOQPPVLYQ-----QPKP-----GALIRPQV 410
QY 404 TGETSGAICLPSTVPKPSVQSVAVPTSVTVTPGKPLNTVTTLKPSLGCASSTPSNEPLKKAENSA 463
Db 411 TLTQT-----PVALQRP-NRIMLTTPQDQL----- 437
QY 464 LPSLFHVQOQSGGNEKQVTTTSHSSTLTQKCGQKTMVNTIIPTSQPPASILKQITL 523
Db 438 -----NPLQPPVYKPAVL 451
QY 524 PGNKIL---SLQASPTQKNRIKENTSCFRDE 552
Db 452 PGTKALSAVSAQAAAKKMLKEPGGGSFRDD 483

```

RESULT 3
 US-08-646-715-16
 Sequence 16, Application US/08646715
 Patent No. 5637686
 GENERAL INFORMATION:
 APPLICANT: Tjian, Robert
 APPLICANT: Comal, Lucio
 APPLICANT: Dynlacht, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoko
 APPLICANT: Wang, Edith
 APPLICANT: Weinzierl, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS.
 TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TARS AND METHODS OF USE
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

```

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-16

Query Match          19.8%; Score 547; DB 1; Length 737;
Best Local Similarity 30.2%; Pred. No. 1.4e-37;
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21;

QY 13 PPRVSSG-----PRIPAPQIVAVKAPNTTIOFPANLQLPFGTVLIRKNSGFLM 61
DB 68 PRATTSIGRATLPFTVLAPRLPQ-----PQNTNIO---NFQLPGMVLVRSNGQL 118
QY 62 LVBPQCVTR-----AETSNITSRAVPANPQTVKICIVPNSSQILIKVAVTPKKL 115
DB 119 MI-PQQLAQMQAQAHQAPQTTAPRAPTPSAPVQISTVQAGTPILNR-QVTP---- 172
QY 116 AQIGTEVTVTPKPSVQSAVPTSVTVTPGKPLNTVT--TLKPSISGASS-----TPS 168
DB 173 -----TIIKQV---SQKQTVQPSATIQRSBGVQPVQLVGGAAQTASLGTATVQGTGTPQ 225
QY 169 NE-PNKAENSAVQINLSPTMLEENVKCKNFMLMLIKLACSGSGSPDMGQNVKLEVO 227
DB 226 RTVPGATTTSSAATE-----TMENVAKCKNFSTLKLKASSGQKSETANVKEIVQNL 279
QY 228 LDKAIEAEFTKRIYVLEKSSPOPHLVPLKKSVALROLIPNSOSTIOQCVQO-----TS 283
DB 280 LDKIEAEFTSRILYRLNLSPPQYLVPLFKRSLLPALROLTPDSAAFIQSSQOQPPPTS 339
QY 284 SDNVIACTTNTVTPVTTVSSQSEKSIYSGATAPRTVSQTLNPLAGPVGARAGV 343
DB 340 Q-----ATTALTAVLVSSVQRTAKTATYTSALQPPVLSL----- 376
QY 344 VTLHSVGPATAGTGTAGTLLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLP 403
DB 377 -----TQPTQVGVCKQGPPLVLIQ-----QPKP-----GALIRPPOV 410
QY 404 TRGETGAALCLPSVKRVVFCMDHICKPVYIGTPVQIKLAQPGVLSQAPACIPGSSSK 463
DB 411 TLQQT-----PVALROPH-NRIMLTTPQOIOL----- 437
QY 464 LFSLFHVQOPSGNKEQVTTISHSSTLTIOKCGQKTPMPVNTLIPTSOPFASILKQITL 523
DB 438 -----NPLQDPVVPVVKRAVL 451

```

```

QY 524 PGNKIL---SLQASPTQKNRIKENVTSCFRDE 552
DB 452 PGTALSAVSAQAAAAQKNKIKLEPGCGSFRDD 483

RESULT 4
US-08-188-582-2
Sequence 2, Application US/08188582
Patent No. 5534410
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-2

Query Match          9.3%; Score 256; DB 1; Length 921;
Best Local Similarity 22.9%; Pred. No. 5.5e-13;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 PQIYAVKAPNTTIOFPANLQLPFGTVLIRKNSGFLM-----VSPQTVRAEFTS 76
DB 133 PQSSTILSTLNTGOTPA-----LLVKTNDNGQLRVGTTGPPVTVQITNTSNN 184
QY 77 NITSRAVPANPQTVKICTVFNSSQ-----LIKKVAVTPVKKLAQIGTVVTVVP 127
DB 185 NITSTNHPITTO--IRLOTVPAAASMTTATSNIIYVASSGYANSSQPPHQLQMAQ 243
QY 128 KPSSVQSAVPTSVTVTPGKPLNTVTTLKPSISGASSTPSNEPNLKAENSAVQINLSP 187
DB 244 APQLPQIQTQITPAQSSQOQVANNVSSAGGTATAVASTTA-----ATT 287
QY 188 TMLNVR-KCKNFMLMLIKLACSGSGSPDMGQNVKLEVO LDKAIEAEFTKRIYVLEK 246
DB 288 TQOGNTEKCKRFLANIEL--STREPRKPVKENVTTIQELVANNVDEEFCDRLERLIN 345

```


FILING DATE: 12-APR-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57503-1/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-046-585-5

Query Match 8.0%; Score 221; DB 1; Length 2035;

Best Local Similarity 23.0%; Pred. No. 1.6e-09;
Matches 131; Conservative 74; Mismatches 211; Indels 154; Gaps 27;

QY 10 VSAPPKVSSGRLPAQIVAVKAPNTTIOFPANLQLPFGTVLIKNSGMLVSPQQT 69
 DB 537 IGSSPQSGMAALAAAATOKIPSSA--PVLSPAGTTIVKT-----MAVTPGTTT 588
 QY 70 TRAEITNITSRPVAVPOTVNTICIVNSSSOLIKKVAVPVKKLAQIGTTV----- 122
 DB 589 LPA--TVKAVSSPVAWSNPAT-----RMLKTRA-----AOGVSASANTTS 628
 QY 123 ---VTVPKPSVQSAVAVPTSVTVTPGKPLNTVTLK-PSSL-GASSTPSNEPNTL---- 173
 DB 629 TRPIITVHKSGTV-TVAQQAQVTVTVGVTKTITLVKSPISVPGSALISNLGKMSV 687
 QY 174 --KAENSAVAQINISPTMLENVKCKNFL--AMLIKACSGSOSP-----EMGQNVK 221
 DB 688 QTRPVQTSVAVTGQASTGSPVTOIOTKGPLPAGTILKLVTSADGKPTTIITTTQASGAGTK 747
 QY 222 KLVEQLDAKIEAEFEFRKLYVELKSSPOPHLVFLKSSVALROLPLNSQSFIOQCVOQ 281
 DB 748 PTILGI-----SSVSPSTT---KPGTTTIKTITPMSAIIITQACATG 785
 QY 282 TSSDMVATCTTIVTT-----SPVVT-----TVSSSOSEKSIIVSGA-----TA 321
 DB 786 VTSSPGIKSPIITITTKVMTSGTGAPAKIITAVPKIATGHGQGVTVVLKGAAGQGTI 845
 QY 322 PRT-----VSQTLNPLAGPVGAK--AGVTLHSVGPPTAATGTTAGTGLQTS 368
 DB 846 LRTVPMGAVRLVTPVTVSAVKAAPTTLVVKGTGTGTVTGIVST--SLAGAGHSTS 903
 QY 369 KPLVTSVA--NTVTVSLOPEKPVSGTAVTSLPATVFGTSGAALCLPSVKPVSCFW 426
 DB 904 ASLATPTITLTATLSSQ---VINPALTIVSAQOTTLTAAGGLTPTTITMOPV----- 954
 QY 427 DHICKPVIGTPVOIKL-AOPGPVLSQPA-GIPTSSSKOLFSLFHVVOQPSGNEKQVTT 484
 DB 955 -----SQPTQVLTILTASGVAEQPVHLPVS-----ILASP-----TT 987
 QY 485 ISHSSTITIOKCGOKTIPVNTIITPSQFP 514
 DB 988 EOPTATVTLNDSQGDVQPGTIVTLVCSNPP 1017

RESULT 7
 US-08-393-703-5
 Sequence 5, Application US/08393703
 Patent No. 558239
 GENERAL INFORMATION:
 APPLICANT: Lamarco, Kelly
 APPLICANT: Wilson, Angus
 APPLICANT: Heit, Winship
 TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:

TITLE OF INVENTION: HOST CELL FACTOR
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/393,703
 FILING DATE: 24-FEB-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57503-2/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-393-703-5

Query Match 8.0%; Score 221; DB 1; Length 2035;

Best Local Similarity 23.0%; Pred. No. 1.6e-09;
Matches 131; Conservative 74; Mismatches 211; Indels 154; Gaps 27;

QY 10 VSAPPKVSSGRLPAQIVAVKAPNTTIOFPANLQLPFGTVLIKNSGMLVSPQQT 69
 DB 537 IGSSPQSGMAALAAAATOKIPSSA--PVLSPAGTTIVKT-----MAVTPGTTT 588
 QY 70 TRAEITNITSRPVAVPOTVNTICIVNSSSOLIKKVAVPVKKLAQIGTTV----- 122
 DB 589 LPA--TVKAVSSPVAWSNPAT-----RMLKTRA-----AOGVSASANTTS 628
 QY 123 ---VTVPKPSVQSAVAVPTSVTVTPGKPLNTVTLK-PSSL-GASSTPSNEPNTL---- 173
 DB 629 TRPIITVHKSGTV-TVAQQAQVTVTVGVTKTITLVKSPISVPGSALISNLGKMSV 687
 QY 174 --KAENSAVAQINISPTMLENVKCKNFL--AMLIKACSGSOSP-----EMGQNVK 221
 DB 688 QTRPVQTSVAVTGQASTGSPVTOIOTKGPLPAGTILKLVTSADGKPTTIITTTQASGAGTK 747
 QY 222 KLVEQLDAKIEAEFEFRKLYVELKSSPOPHLVFLKSSVALROLPLNSQSFIOQCVOQ 281
 DB 748 PTILGI-----SSVSPSTT---KPGTTTIKTITPMSAIIITQACATG 785
 QY 282 TSSDMVATCTTIVTT-----SPVVT-----TVSSSOSEKSIIVSGA-----TA 321
 DB 786 VTSSPGIKSPIITITTKVMTSGTGAPAKIITAVPKIATGHGQGVTVVLKGAAGQGTI 845
 QY 322 PRT-----VSQTLNPLAGPVGAK--AGVTLHSVGPPTAATGTTAGTGLQTS 368
 DB 846 LRTVPMGAVRLVTPVTVSAVKAAPTTLVVKGTGTGTVTGIVST--SLAGAGHSTS 903
 QY 369 KPLVTSVA--NTVTVSLOPEKPVSGTAVTSLPATVFGTSGAALCLPSVKPVSCFW 426
 DB 904 ASLATPTITLTATLSSQ---VINPALTIVSAQOTTLTAAGGLTPTTITMOPV----- 954
 QY 427 DHICKPVIGTPVOIKL-AOPGPVLSQPA-GIPTSSSKOLFSLFHVVOQPSGNEKQVTT 484

Db 955 -----SQPQVTLITAPSGVEAQPVDLPVS-----ILASP-----TT 987

QY 485 ISHSTLTIOKCGQKTMPTVTIIPISQFPP 514

Db 988 EQPATVTTIADSGQGDVQPGTVTLVCSNP 1017

RESULT 8

PCT-US93-11721-5

Sequence 5, Application PC/TUS9311721

GENERAL INFORMATION:

APPLICANT: Lamaco, Kelly

APPLICANT: Willson, Angus

APPLICANT: Heert, Winship

TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:

TITLE OF INVENTION: HOST CELL FACTOR

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/11721

FILING DATE: 03-DEC-1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: FP-57503-1/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 396-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

PCT-US93-11721-5

Query Match 8.0%; Score 221; DB 5; Length 2035;

Best Local Similarity 23.0%; Pred. No. 1.6e-09;

Matches 131; Conservative 74; Mismatches 211; Indels 154; Gaps 27;

QY 10 VSAPPKSSGPRLPAPQIYAVKAPNTTTIOFPANLQLPQGTVLKNSGRLMVSQOQY 69

Db 537 IGSSTQMSGMALAAAAATOKIPSSA---PVLSPVPGTITVKI-----NAVTEGITT 588

QY 70 TRAEFTSNITSRPAPVPANQVQVKTIVPNSSQLIKKVAVTVPYKLAQIGTV-----122

Db 589 LPA--TVKVAASPVAVSNPAT-----RMLKTA-----AQGTSVSATMNS 628

QY 123 ---VTVTPKSSVQAVATSVYVTPGKPLMTVTTLK-PSL-CASSTPSNEPNL-----173

Db 639 TRPITTVHSGTV-TVAAQAAQVTVTVGVTIKTITLVKSPISVPGSALISNLGKVMYV 687

QY 174 --KAENSAVQINTLSPTEMLNKCKNFL--AMLIKLACSGOSP-----EMGQNVK 221

Db 688 QKRPVQTSVAVTQASTGPTVTOIQTGKPLPAGTITLKVTSADCKPTTITTTTQASGAGTK 747

QY 222 KLVEOLDLAKIAEETRKLVEELKSSPOPHLVPELKKSVALLQLLPNSQSEFIQOCVQO 281

Db 748 PTLIGI-----SSVSPSTT---KPGTTIIKIIPSAITTOAGANG 785

QY 282 TSDMVAIATCTVTT-----SPVVT-----TVSSSSQSEKSIIVSGA-----TA 321

Db 786 VTSSPGIKSPIITITTKVMTSGTGAPAKITFAVKIATGCGQGVQVYLKGAQCQGTI 845

QY 322 PRT-----VSQVTLNPLAGPVAK--AGVTLHSGVPAATNGTGTAGTGLQTS 368

Db 846 LRTVPMGCVLRVTPVYSAKPAVTTLVKGTGTGTGTGTGT--SLAGAGHSTS 903

QY 369 KPLVTSVA--NTVTVSLQPEKPVSGTAATLSLPAVTFGETSGAICLPSPVPVPSFCW 426

Db 904 ASLATPITTTGTTATLSSQ---VINFTATVSAQVTLNAGILPTITMQPV-----954

QY 427 DHICKPVITGPVQIKL-AQGPVLSQPA-GIPGSSSKQLSLFHVYQDPSGNGEKVTT 484

Db 955 -----SQPQVTLITAPSGVEAQPVDLPVS-----ILASP-----TT 987

QY 485 ISHSTLTIOKCGQKTMPTVTIIPISQFPP 514

Db 988 EQPATVTTIADSGQGDVQPGTVTLVCSNP 1017

RESULT 9

US-09-579-181-2

Sequence 2, Application US/09579181

Patent No. 6365372

GENERAL INFORMATION:

APPLICANT: Chivlia, John

APPLICANT: Yaciuk, Peter

TITLE OF INVENTION: SNE2 Related CBP Activator Protein (SRCAP)

FILE REFERENCE: 16153-4247

CURRENT APPLICATION NUMBER: US/09/579,181

CURRENT FILING DATE: 2000-05-25

PRIOR APPLICATION NUMBER: 60/136,620

PRIOR FILING DATE: 1999-05-27

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 2

LENGTH: 2972

TYPE: PRT

ORGANISM: Human

US-09-579-181-2

Query Match 7.1%; Score 197; DB 4; Length 2972;

Best Local Similarity 23.7%; Pred. No. 2.9e-07;

Matches 140; Conservative 75; Mismatches 240; Indels 136; Gaps 32;

QY 2 TLVTKVAPASAPPKVSSG-----PRLPAPQIYAVKAPNTTTIOFPANLQ-- 46

Db 1063 TPVPLAPAPRPP--SSGLEPAVLNPRPTLTGKRLPPTLTGTARAPMPTPLVPLKLKH 1120

QY 47 --PPGTVLKNSGRLMVSPOQVTRAETTSNITSRPAPVPANQVQVKTIVPNSSQOLI 104

Db 1121 SPSEPVASAPGAAPLTISSPLHP-----SSLPGASSPM-----PINSPLIAS 1166

QY 105 KKAIVTPYKLAQIGTIVTVTPKSSVQAVPTSY-VTVTPGKP-LMTVT-TLKPS 160

Db 1167 PVSSVSVPLSSSLPISVPTTLTPAPSA--PLTIPISAPLVLSASGPAALLTSVPPLP 1225

QY 161 LGASSTPSNEPNLKAENSAVQINTL--SPTEMLNKCKNFLMLIKLACSGOSPSEM 218

Db 1226 PALPGPSLQPSGASPSASALITGLATAPSLSSQTPGPHL-----LAFSTSHVGLNS 1280

QY 219 NVKRLVEQ-LIDAKIAEETRKLVEELKSSPOPH-----LVPLKSSVVALRQLLPNS 271

Db 1281 TVAPACSPVLVPASALASPE-----PSAPRPARQASLAPASASGALATPLAPMA 1332

QY 272 QSFIOQCVQOQSDMVAIATCTVTTSPVYTTTVSSQSEKSIIVSGAT-----320

Db 1333 AP--QTAIILAPSPAPPLAP-LPVLAPSPGAPVLAASSQTPVPMAPSPFGTSLASAPV 1389

QY 321 -APRTV-----SVQTL--NPLAGPVAKAGVTVLHSGVPTNA--TGCT-----TA 365

Db 1390 PATTPLABSSOTMTLPAPVPSLPSPASTQTL-ALAPALAPTLGSSPSQTLSTGTGNP 1448
Oy 366 QTSKPLVTSVANTVTTSLOPEKPVYSGTAVTSL-PAVTEGETSGAIC-LPSVPRVVS 423
Db 1449 QGFFP-----TQTLSTLPASSLVPTPAQTLSTLAPGPPGPTQTLSTLAPAPPLAP--- 1497
Oy 424 FCMDDHICKPVYIGTPV-----QIKLAQGPVLS---QAPGICPT-GSSSKQLPSLFH 469
Db 1498 -----ASPYPAPAHMTLTLAPASSASLAPASVQTLTSLPAPVPTLGPMAAQTLALAP 1551
Oy 470 V-VOQPSGNEKQVTTISHSTLTIOKCGOKTAPVNTI--IPTSOPEPASI 517
Db 1552 ASTQSPA-----SQASSLVASASGAPLPVTWVKSLPVSKEDEPDTL 1592

RESULT 10
US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 635372
; GENERAL INFORMATION:
; APPLICANT: Chiviva, John
; APPLICANT: Yaculik, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247 US/09/579,181
; CURRENT APPLICATION NUMBER: US/09/579,181
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3118
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-1

Query Match 7.1%; Score 197; DB 4; Length 3118;
Best Local Similarity 23.7%; Pred. No. 3, 2e-07;
Matches 140; Conservative 75; Mismatches 240; Indels 136; Gaps 32;

Oy 2 TLVTKAPVAPAPKVS-----PRLPAPQIVAVKAPNTTITQFANQL-- 46
Db 1209 TTPVPLAPAPRRP--SSGLPAVINPRLPTLPGRLPTLTGTAAPMTPTLVNPLKLVH 1266
Oy 47 --PPTGLIKSNGLMLVSPQQTVAETTSNITSRPAVAPQVTKICTVNSSQLT 104
Db 1267 SPSEVASASAPGAAPLTISPLHP-----SSLPGASSPM-----PINSSPLAS 1312
Oy 105 KKYAVTPVKKLAOIGTTVTTPKSSVQSAVPTSV-VTVTPGKP-LNFTVT-TLKPS 160
Db 1313 PVSTVSVPLSSSLPISVPTTLPAPASA-PLTIPISAPLTVSASGALLTSVPPLAPV 1371
Oy 161 LGASSTPSPNEPNLKAENSAVOINL--SPTMLENVKKCKKFLMLIKLACSSGSPBMQ 218
Db 1372 PAAPGPPSLQPSASPSASLTLGLATAPLSLSSQTPGHL-----LAPTSSHVPLNS 1426
Oy 219 NYAKLVEQ-LLDKIAEEETRLKLYELKSSPOPH-----LVFLKKSVALROLPLNS 271
Db 1427 TVAPACSPVLVPASALASPE-----PSAPMPAPAQASLAPASASALATPLAPMA 1478
Oy 272 QSFITQCVQOOTSDMVIAFCTTVTTSPPVTTVSSOSEKSIIVSGAT----- 320
Db 1479 AP--QTAALAPSPAPPLAP-LPVLAPSPGAPVLAASSQTPVPVMASSPTSGTSLASAPV 1535
Oy 321 -APRTV-----SVQTL--NPLAGPVGAKAGVTLHSYGPTAA--TGCT--TAGTGL 365
Db 1536 PATTPLABSSOTMTLPAPVPSLPSPASTQTL-ALAPALAPTLGSSPSQTLSTGTGNP 1594
Oy 366 QTSKPLVTSVANTVTTSLOPEKPVYSGTAVTSL-PAVTEGETSGAIC-LPSVPRVVS 423
Db 1595 QGFFP-----TQTLSTLPASSLVPTPAQTLSTLAPGPPGPTQTLSTLAPAPPLAP--- 1643
Oy 424 FCMDDHICKPVYIGTPV-----QIKLAQGPVLS---QAPGICPT-GSSSKQLPSLFH 469

Db 1644 -----ASPYPAPAHMTLTLAPASSASLAPASVQTLTSLPAPVPTLGPMAAQTLALAP 1697
Oy 470 V-VOQPSGNEKQVTTISHSTLTIOKCGOKTAPVNTI--IPTSOPEPASI 517
Db 1698 ASTQSPA-----SQASSLVASASGAPLPVTWVKSLPVSKEDEPDTL 1738

RESULT 11
US-08-325-267A-4
; Sequence 4, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: MATARI, JUNJI
; APPLICANT: TAKATA, YOSHIRO
; APPLICANT: OGAWA, MASAHITO
; APPLICANT: PENITILLA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 862 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-267A-4

Query Match 7.0%; Score 193; DB 1; Length 862;
Best Local Similarity 20.1%; Pred. No. 1e-07;
Matches 120; Conservative 88; Mismatches 216; Indels 174; Gaps 24;

Oy 11 SAPPKVSGRLPAPQIVAVKAPNTTITQF-----PANLQLPPTGLIKSNGL----- 60
Db 199 SLPNIEGTYVMAGYGYPPKVVYSNAVSWGTLPISVTLPDGTVSDDPEGVYSEDDDL 258
Oy 61 -----MLVSPQQTVAET-----TSNITSRPAVAPNPOTVACIVPNS 99
Db 259 SQSNCTVPPDSNNAVSTTTTTEPMTGTFTSTEMTIVTGINGVFPD-ETVIVIRPTT 317
Oy 100 SSQILIKKVAVTPVKKLAOIGTTVTTPKSSVQSAVPTSVTVTPGKPLN-TVTTLK- 157

```

Db 318 AS-----TITTEPWTGFTSTSTMTTGTGNGOPTDETVIVNT 359
QY 158 PSSLGASSTPSNEP-----NKAENSAVAOINLSPTMLENKKCNFLAMLIKILACSGS 211
Db 360 PTSEGLVTT-TTEPMTGFTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 416
QY 212 GSEPMGNKVKIIEQLDLAKIAEETRKLYELKSSPOPHLVPLKKSVALROLPLNS 271
Db 417 SSGQITTSI-----TSSRPITTFPPS-----NG 440
QY 272 GSEFQOCVQOTSSDMVIANCTTV-TTSPVPTTVSS-----SOSKSIIV----- 316
Db 441 TSVT-----SSVYSSSVTSLSFTSSPVYSSSVTSSTTTSTSTSTSTSTSTSTSTST 493
QY 317 --SGATAPRTVSQVTLNPLAGPVGAKAGVYTLHVGPTATGATGTGTLTOSKPLVS 374
Db 494 STSGSESESTSS-----AGSVSSSSFTSSSSKSPYSS-----SLPLVNS 535
QY 375 VANITYTVSLOPEKPVVSGTAVTSLPVAFTGEGTSGAICLPSPVSPFCMDHIC----- 430
Db 536 ATTSOETAS-----SLPPTTTKTSEOTTLV-----TVSCSHVCTESI 575
QY 431 KPVIGPVOIKLAOPGVLPSPAGIPTGSSSKQLPSLFHVVOOPSGGNEKQVTTISHST 490
Db 576 SPAVSTAVTV-----SGVTETTTWCPISTTETTKOTKCTTQTETTKOTV 625
QY 491 LTIQKCGKTMPEVNTIPTSOFPASILKQITLPGNKILSLQASPTOKNRIRENTSC 548
Db 626 VTISSC-ESDVCSTKASPAVSTSTATNGVTETTTWCPISTTESRQOTTLVTVTSC 682

RESULT 12
US-08-362-525-22
Sequence 22, Application US/08362525
Patent No. 6027910
GENERAL INFORMATION:
APPLICANT: KUS, FRANCISCUS M.
APPLICANT: SCHREUDER, MARTEN P.
APPLICANT: TOSCHKA, HOLSTER Y.
APPLICANT: VERRIPS, CORNELIS T.
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESS: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,525
FILING DATE: 04-JAN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92202080.5
FILING DATE: 08-JUL-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92203899.7
FILING DATE: 14-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/01763
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773

```

```

REFERENCE/DOCKET NUMBER: 213289/T7020(V)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-525-22

Query Match 6.9%; Score 189; DB 3; Length 894;
Best Local Similarity 20.7%; Pred. No. 2,3e-07;
Matches 118; Conservative 84; Mismatches 217; Indels 152; Gaps 24;

QY 22 LPAQIVAAKAPRTTIOFPANIQLPVGLIKNSGRLMYSPOQ-----TYTRAE 73
Db 265 VDPSPNVAVSTTTT-----EPWTGFTSTSTMTTGTGNGVPTDETVIVIRTP 315
QY 74 TTSNISRPAVP-----ANQVAKICTVPMSSOLIKKAV---TPVKLAQIGTVVTV 126
Db 316 TSBGLSTTTEPWTGFTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 371
QY 127 PKPSSVQSAVPSTSVTVTPGKPLN-TVYTLK-PSSLGASSTPSNEP-----NKAENS 178
Db 372 PWTGFTSTSTMTTGTGNGOPTDETVIVIRTPSEGLVTT-TTEPMTGFTSTSTST 430
QY 179 AAVQIULSPTMLENKKCNFLAMLIKILACSGSQSPMGONKVKIIEQLDLAKIAEET 238
Db 431 TVTGNGLPTD-ETVIVVNT-PTTAISSLSLSSSGQITTSI----- 470
QY 239 RKLVELKSSPOPHLVPLFKKSVALROLPLNSQSEFIQOCVOQOTSSDMVIANCTTV-TT 297
Db 471 -----TSSRPITTFPPS-----NGTSVI-----SSVYSSSVTSLSFT 505
QY 298 SPVYTTTVSS-----SOSKSIIV-----SGATAPRTVSQVTLNPLAGVGAKA 341
Db 506 SPVYSSSVTSLSFTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 558
QY 342 GVTTLHVGPTATGATGTGTLTOSKPLMVSANVTYTVSLOPEKPVVSGTAVTSLP 401
Db 559 FTSSSESKSPYSS-----SLPLVNSATTSOETAS-----SLP 592
QY 402 AVTFGTSGAICLPSPKPVVSCFMDHIC-----KPYIGTPOIKLAOPGVLPSPAGIPT 457
Db 593 PATTTSTSQTLV-----TVSCSHVCTESISPAVSTAVTV-----SGVT 637
QY 458 GSSSKQLPSLFHVVOOPSGGNEKQVTTISHSTLTIOKCGKTMPEVNTIPTSOFPASI 517
Db 638 EYTWCPITSTETTKOTKCTTETOTETTKOTVTVTSSC-ESDVCSTKASPAVSTSTAT 696
QY 518 LKQITLPGNKILSLQASPTOKNRIRENTSC 548
Db 697 INCVTETTTWCPISTTESRQOTTLVTVTSC 727

RESULT 13
US-08-971-692-15
Sequence 15, Application US/08971692
Patent No. 611417
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Immobilized proteins with specific binding
TITLE OF INVENTION: capacitacs and their use in processes and products.
NUMBER OF SEQUENCES: 40
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/971,692
 FILING DATE:
 CLASSIFICATION: 435
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 894 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-971-692-15

Query Match 6.9%; Score 189; DB 3; Length 894;
 Best Local Similarity 20.7%; Pred. No. 2,3e-07;
 Matches 118; Conservative 84; Mismatches 217; Indels 152; Gaps 24;

QY 22 LPAPQIVAKKAPNTTIOFPANLQLPQTVLIKNSGPIMLVSPQQ-----TVTRAE 73
 DB 265 VPDSNNVAVSTTTT-----EPMTGFTSTSTEMTVTGTNGVPTDEVIYVIRP 315
 QY 74 TTSNISRPAVP-----ANPQTVKICVPPNSSQLIKKVAV--TPVKLAQIGTVTVTV 126
 DB 316 TSEGLISTTEPMTGFTSTSTSTEVTTITGTNGOPTDEVIYIRPTSE---GLISTTE 371
 QY 127 PKSSVQAVAVPTSVTVTPKPLN-TVTTLK-PSSLGASTPSNEP-----NLKAENS 178
 DB 372 PMTGTFTSTSTEMTVTGTNGOPTDEVIYIRPTSEGLVTT-TTEPMTGFTSTSTEMS 430
 QY 179 AAVQINISPTMLENVKCKNFIAMIKIACSGSQSPKMGQVKKIQLDAKIEAEFT 238
 DB 431 TVTGTNGLPTD-ETVIYVKT-PTTAISSLSLSSSSGQITSI-----470
 QY 239 KKLVEELKSSFPQHLVPLKKSVALNQLPNSQSFIOQVQOQSSDMVIACTTV-TT 297
 DB 471 -----TSSRPITTFEYPS-----NGTSVI-----SSSVISSVTSLSPTS 505
 QY 298 SPVYTTVSS-----SSEKSIIV-----SGATRPVSVQVLNLAGPVGKA 341
 DB 506 SPVSSSVSSVST 558
 QY 342 GVVTLHSGVPAATGCTAGTGLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLP 401
 DB 559 FISESSKSPYSS-----SLPLVTSATSETMS-----SLP 592
 QY 402 AVTFGETSGAICLPKPVVSCWDHIC---KPVIGTPVQIKLAGPVLSOPAGIPT 457
 DB 593 PATTKISSEQTLV-----TVTSCESHVCTESIPAVSTATVTV-----SGVTT 637
 QY 458 GSSSKQLEFLFHVQOPSGNEKQVTTISHSTLTLOKCGKTPVNTIPTSOPPAISI 517
 DB 638 EYTWCPISSTETTKQKGTTEQTTETTKQTVVTIISSC-ESDVCSTKASPAIVSTAT 696
 QY 518 LKQITLPGNKILSLQASPTQKRIKENVTSC 548
 DB 697 INGVTTETWCPISTETSKQOTLIVTVTSC 727

RESULT 14
 US-08-325-267A-2
 Sequence 2, Application US/08325267A
 Patent No. 5585271
 GENERAL INFORMATION:
 APPLICANT: MATSURI, JUNJI
 APPLICANT: TAKATA, YOSHIHIRO
 APPLICANT: OGAWA, MASAHIRO
 APPLICANT: PENTTILA, MERJA
 APPLICANT: ONNELA, MAIJA-LEENA
 APPLICANT: KERANEN, SIRKA
 TITLE OF INVENTION: YEAST AGGLOUTINATION GENES AND YEAST
 TITLE OF INVENTION: CONTAINING THEM
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
 STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400

CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/325,267A
 FILING DATE: 18-NOV-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP PCT/JP94/00290
 FILING DATE: 24-FEB-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 38871/1993
 FILING DATE: 26-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: OBION, NORMAN F
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 2589-023-0XPCF
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-413-3000
 TELEFAX: 703-413-2220
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1537 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-325-267A-2

Query Match 6.5%; Score 179.5; DB 1; Length 1537;
 Best Local Similarity 20.5%; Pred. No. 3.2e-06;
 Matches 109; Conservative 82; Mismatches 195; Indels 147; Gaps 22;

QY 50 TVLI---KNSGPIMLVSPQOTVIRAEITSNITSNPANPAP--QTVKICVPPNSSQLI 104
 DB 938 TVIYIRPTSEGLISTTEPMTGFTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 993
 QY 105 KKVAVTPVKLAQIGTVTVTPKSSVQAVAVPTSVTVTPKPLN-TVTTLK-PSSLG 162
 DB 994 -----GLISTTEPMTGFTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1039
 QY 163 ASSTPSNEP-----NLKAENSAVQINLSPTMLENVKCKNFIAMIKIACSGSQSP 216
 DB 1040 LVTT-TTEPMTGFTST 1096
 QY 217 GQNVKQIVLEQLDKIEAEFTKRIYVELKSSPOHVLVPLKKSVALROLPNQSFIO 276
 DB 1097 TTSI-----TSSRPITTFEYPS-----NGTSVI- 1119
 QY 277 QCVQQTSSDMVIACTTV-TTSPVYTTVSS-----SSEKSIIV-----SGA 319
 DB 1120 -----SSSVISSVTSLSLFFISSPVISSVTSSTSTSTSTSTSTSTSTSTST 1173
 QY 320 TAPRTVSVQTLNPLAGPVGAKAGVTVLHSGVPTAATGTTAGTGLQTSKPLVTSVANTV 379
 DB 1174 SESSETSS-----AGSVSSSSSFISSESKSPYSS-----SLPLVTSATTSQ 1215
 QY 380 TTVSLQPEKPVVSGAVAVLSPAVTFGETSGAICLPKPVVSCWDHIC---KPVIG 435
 DB 1216 ETAS-----SLPPATTTTTSQTLIV-----TVTSCESHVCTESIPAVI 1255
 QY 436 TPVQIKLAQPPVLSOPAGIPTGSSKQLEFLFHVQOPSGNEKQVTTISHSTLTLOK 495
 DB 1256 STAVTV-----SGVTTETWCPISTETTKQKGTTEQTTETTKQTVVTIIS 1305
 QY 496 CGQKTPVNTIPTSOPPAISILKQITLPGNKILSLQASPTQKRIKENVTSC 548

Db 1306 C-ESDVCKSTAPSAIVSTATINGVTETWCPISTESROOTLIVTSC 1357

Search completed: February 16, 2003, 22:02:41
Job time: 29.1842 secs

RESULT 15
US-08-700-651-5

Sequence 5, Application US/08700651B

Patent No. 6015882

GENERAL INFORMATION:

APPLICANT: PETERSEN, CAROLYN

APPLICANT: LEECH, JAMES

APPLICANT: NELSON, RICHARD, C.

APPLICANT: GUT, JIRI

TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs

TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum

FILE REFERENCE: 480.19-4(HV)

CURRENT APPLICATION NUMBER: US/08/700,651B

EARLIER FILING DATE: 1997-08-14

EARLIER APPLICATION NUMBER: 08/415,751

NUMBER OF SEQ ID NOS: 15

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 1721

TYPE: PRT

ORGANISM: Cryptosporidium parvum

US-08-700-651-5

Query Match

Best local similarity 20.9%; Pred. No. 1e-05;

Matches 118; Conservative 65; Mismatches 237; Indels 145; Gaps 23;

QY 35 TTTTIOEPANILPPGTVLIRKNSGPMVSPQQTVAETTSNITSRAVAPNPQTVKIC 94
Db 286 TTTTITT 345
QY 95 TVNSSSQLKKAAVPPVKLAIGTTVTVTKPSSVQSAVPTSVYTPGKPLNTV 154
Db 346 TTTTITT 398
QY 155 TLKPSLIGASSTPS-----NEPLKAENSA-----VOINLSPMLEN-- 192
Db 399 TTTTITT 458
QY 193 -----VKKCNFLAMLIKLAGSGSQSPMGONVAKLVEQLDAKIE--AEFTKRLY 242
Db 459 THVRFREKYADVONTISVRCRKAGKLEFP-----DRSLDEFTIPVACHNSCII 508
QY 243 VELKSSPQPLVPLFKSVVALRQLPNSQSF-----IQQCV-----QOTSSDVIA 289
Db 509 VGVSGDGKIHVSPIGSDVSLISAPIOPSELEFNEVYCDCTAKYGAHSGVOTSDPVT 568
QY 290 TC---TTTVTSPVVTTVSSOSEKSIIVSGATAPRTVSVOPLNPLAGVGAAGVTL 346
Db 569 TTKKPTTTTGAAGOPTTTTTSKSPK--TTTTTKATTTTTTLNP-----IITF 616
QY 347 HSGVPTAAGGTAGGLQTSKPLVTSVANVTVVSLOPEKPV-----SGTAVTSL 400
Db 617 TTKKPTTTT--TTKVPG-----KP--PLATTTTTL-----KPIVTTTTKATTTTTT 661
QY 401 PAVTFETSGAALCLPVSVPVSVFCHDICKPIY-----TPVOIKLAQPGVLSQPA 456
Db 662 PTTT--TTTKRDEMTTTTPL-----PDIGIEITPIPIE----- 694
QY 457 TCGSSKQLSFLHVOOPSGGNEKQVTTISHSSTLTIQKCGKTMPTVNTIIPTSOPPAS 516
Db 695 -----KMLDKYTMITDYNSG-----LLDSNDPPIGSOAGQ-----IADTSNLEPV 738
QY 517 LKQITLPGNKIISLQASPTQKNRI 541
Db 739 THKSTGLPIDPMVGLPDPKSNLV 763

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 10.8745 Seconds

(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GTLVTKVAPVSAAPKVSQGP.....ASPTOKRIKENTSCFRDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gaepct 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCr_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCrUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	521	18.9	1023	9 US-09-893-519A-14	Sequence 14, Appl
2	220	8.0	1367	10 US-09-801-368-108	Sequence 108, App
3	219	7.9	5179	9 US-10-025-380-1068	Sequence 1068, App
4	219	7.9	5179	10 US-09-922-217-1068	Sequence 1068, App
5	219	7.9	5179	10 US-09-833-263-1068	Sequence 1068, App
6	195	7.1	1322	10 US-09-801-368-114	Sequence 114, App
7	179.5	6.5	1537	10 US-09-801-368-104	Sequence 104, App
8	177.5	6.4	2665	10 US-09-864-761-34248	Sequence 34248, A
9	176.5	6.4	1601	10 US-09-862-027-40	Sequence 40, Appl
10	174	6.3	1075	10 US-09-801-368-110	Sequence 110, App
11	172	6.2	1169	10 US-09-801-368-106	Sequence 106, App
12	159.5	5.8	1056	9 US-10-161-510-10	Sequence 10, Appl
13	158.5	5.7	2597	10 US-09-905-129-2	Sequence 2, Appl
14	158.5	5.7	2597	10 US-09-905-129-10	Sequence 10, Appl
15	158.5	5.7	2597	10 US-09-905-129-13	Sequence 13, Appl
16	158.5	5.7	2597	10 US-09-991-630-2	Sequence 2, Appl
17	158.5	5.7	2597	10 US-09-991-630-10	Sequence 10, Appl
18	158.5	5.7	2597	10 US-09-991-630-13	Sequence 13, Appl
19	157.5	5.7	941	12 US-10-124-557-14	Sequence 14, Appl

20	157.5	5.7	1022	12 US-10-124-557-84	Sequence 84, Appl
21	157.5	5.7	1038	12 US-10-124-557-74	Sequence 74, Appl
22	157.5	5.7	1049	12 US-10-124-557-58	Sequence 58, Appl
23	157.5	5.7	1140	12 US-10-124-557-104	Sequence 104, Appl
24	157.5	5.7	1270	12 US-10-124-557-44	Sequence 44, Appl
25	157.5	5.7	1311	12 US-10-124-557-42	Sequence 42, Appl
26	157.5	5.7	1313	12 US-10-124-557-142	Sequence 142, App
27	157.5	5.7	1314	12 US-10-124-557-50	Sequence 50, Appl
28	157.5	5.7	1320	12 US-10-124-557-46	Sequence 46, Appl
29	157.5	5.7	1354	12 US-10-124-557-48	Sequence 48, Appl
30	157.5	5.7	1361	12 US-10-124-557-40	Sequence 40, Appl
31	157.5	5.7	1363	12 US-10-124-557-52	Sequence 52, Appl
32	157.5	5.7	1404	12 US-10-124-557-2	Sequence 2, Appl
33	157.5	5.7	1404	12 US-10-124-557-62	Sequence 62, Appl
34	157.5	5.7	2587	10 US-09-864-761-36047	Sequence 36047, A
35	156	5.7	2586	10 US-09-905-129-11	Sequence 11, Appl
36	156	5.7	2586	10 US-09-905-129-14	Sequence 14, Appl
37	156	5.7	2586	10 US-09-905-129-11	Sequence 11, Appl
38	156	5.7	2586	10 US-09-991-630-11	Sequence 11, Appl
39	156	5.7	2586	10 US-09-991-630-14	Sequence 14, Appl
40	156	5.7	2587	10 US-09-905-129-16	Sequence 16, Appl
41	156	5.7	2587	10 US-09-991-630-16	Sequence 16, Appl
42	156	5.7	2589	10 US-09-991-630-24	Sequence 24, Appl
43	155	5.6	2005	10 US-09-735-367B-3	Sequence 3, Appl
44	155	5.6	2063	10 US-09-735-367B-2	Sequence 2, Appl
45	153.5	5.6	2828	10 US-09-905-129-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-893-519A-14
Sequence 14, Application US/09893519A
Publication No. US20030027243A1
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G548-US2
CURRENT APPLICATION NUMBER: US/09/893, 519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION: Human Genbank/CAA72189
DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)

US-09-893-519A-14

Query Match 18.9%; Score 521; DB 9; Length 1023;
Best Local Similarity 31.7%; Pred. No. 6,5e-27;
Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;

QY 13 PRVSSG-----PRLPAPQIVAVKAPRTTITQIPANLQIPGTVLKSNSGFLM 61
DB 414 PRTATGIRATLTPVLAAPLPP-----PQNPNTNQ--NQLPFGMLVSENQQL 464
QY 62 LVSPQOTVTR-----AETSNITSRPANPQVAKICIVPNSSSOLIKKAVTPPKYL 115
DB 465 MI-PQALAQMOQAHAQPOQTMAPRATPTSPVQISIVQAPGTITLR-QVTP----- 518
QY 116 AQIGTTVTVTPKPSVQSAVPPTSVVTVPKPLNTVT--TLKPSLGASS-----TPS 168
DB 519 ----TTLIKQV---SOAQTVQPSATLQSRPGVQPOLVLCGAQOTASLGTATVAVQCTPQ 571
QY 169 NE-PNLKAENSAVQINLSSTMLENKKCNFLAMLIKLCSSGSPSEKQNKVLEQL 227
DB 572 RIVPGATTSSAATE-----TMENVKCKNPLSTLIKASSGKOSTETANVKELONL 625
QY 228 LDAKIAEETRLKLYVELKSSQPHLVPELKSVALROLPSQSF10QCVOQ---TS 283
DB 626 LQKIEADETSLRYELKNSPQPIVLPFLKRLPALROLTPDSANFIQSSQOQPPPTIS 665
QY 284 SDNVIACTTTVTSPVVTTVSSSQSEKSIIVSGATAPRTVSVOGLNPLAGPVAKAGV 343
DB 686 Q-----ATALTAVLVSSSVQRTAGKTAIVTSALQPVLSL----- 722
QY 344 VLIHSVGPRAATGTGNGCGLQTSKRLVTSVANTVTVSLOPEKPVVSTAVTSLIPAV 403
DB 723 -----TQPTQVGVGKQGOPTPLVLIQ-----QPKP-----GALIRPQV 756
QY 404 TEGETSGAICLSPVSPVFCMDHICKPVIGTPVQIKLAQPGPVL---SOPAGIPGTS 459
DB 757 TLQGT-----PMVALRQPH-NRIMLTTPQOVNLSSEBARIATATSELVGLITRS 804
QY 460 SSKQFLSFLEHVQ-----QPSGNEKQ--VTTISHSFTLTIQKCGK 499
DB 805 CKDEFLAPLQRLILEIGKKHGITELHPDVSVSVSHATQRLQNLVEK 854

RESULT 2

US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367

TYPE: PR1
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Query Match 8.0%; Score 220; DB 10; Length 1367;
Best Local Similarity 22.5%; Pred. No. 8.7e-07;
Matches 131; Conservative 77; Mismatches 240; Indels 134; Gaps 22;

QY 2 TLVTKAVAPVAPKPVSSGPRLPAP-----QIVAVKAPRTTITQIPANLQIPGTVLIK 54
DB 618 TTESSAPVTSSTTESSAPVPPSSSTTESSAPVPTPSSSTTESSAPVPPSSSTTE 677
QY 55 SNSGPRMLVSPQOT---VTAETSNITSRPANPQVAKICIVPNSSSOLIKKAV-- 109
DB 678 SSSAPVTSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPPSSSTTESSAPV 737
QY 110 -TPVKKLAQIGTVTVTV-----VPRKSSVQS-----VAVPRTVTVTPGKPLNT- 152
DB 738 PTBSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTP 797
QY 153 -----VTLKPSLGASTPSENEPLKAENSAVQINLSPTMLENVKCKNFLA 201
DB 798 SSSSTTESSAPVPTPSSSNTSSAPSPSTPSSSTTESSAPV---PVPSSSTTESSS-- 851
QY 202 MLKLKACSGQS-----PEKQNVKKLYQLDKATEAETTRKLYVELKSSQPHLV 254
DB 852 --APVSSSTTESSAPVPTPSSSNTSSAPSPSTPSSSTTESSSTCTTVTPSSSK----- 904
QY 255 PFLKSVVALROLPSQSF10QCVOQSSDNVIACTTTVTSP---VTTTVSSSQSE 311
DB 905 -----YPSGQT--ETSVSSTTETTVIPKTTTSVTPSTTTTVCSTGTV 949
QY 312 KSIIVSGATAPRTV--SVQTLNPLAGPVAKAGVTLHSVGPRAATGTGNGCGLQTSK 367
DB 950 SAGETTSAGCPKVTVTPTTTTTSVTSSTTTTTCSTGNSAGETTSAGCPKVTIT 1009
QY 368 SKPLVTSVANT---VTVSLOPEKPVVSGTAVTSLIPAV--FGETSGAIC----- 414
DB 1010 TVPCSTSPSETASESTTPTPTVTVSVTVTVTEYSTKPGELTTFVFNKPIPTV 1069
QY 415 -----LPSVKPVVSPFCMDHICKPVIGTPVQIKLAQGPVLSQSGSKQFLSLF 468
DB 1070 LTTIAPPTSVTITVNF-----TPTTTITT---VCS-----TGTNS----- 1101
QY 469 HVVQOPSGNEKQVTTISHSFTLTIQKCGKTPMVTNITPTS 510
DB 1102 -AGETTSAGCPKVTVTTVPCSTGT---GEVTEATTLVTTA 1138

RESULT 3

US-10-025-380-1068
; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skelley, Yeaeli A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14

```
Db 427 TEVTTTGGNGQPPDEF-----VIVIRTPISGLITTTT-----PMTGTFST 470
Qy 57 SGPIMLV---SPQOT-----VTRAETTSNITSRPAPV-----ANPQVKTCTPVSSSOLI 104
Db 471 STEMTVTGNGQPTDEVIIVIRTPISGLITTTTETPTGTFSTSTEVTTITGNGOPT 530
Qy 105 KKVAV---TPVKKLAQIGTVVTVTPRPSVQSAVPTSVVTPGKPLN-TVTTLKRSS 160
Db 531 DEVIIVIRTPISGLITTTTETPTGTFSTSTEVTTITGNGOPTDEVIIVIRPT 586
Qy 161 LGASSTSNENPKAENSAVOI-----NISPTMLENVKCKKFLAMLK---LACSGS 211
Db 587 SEGILTTTETPTGTFSTSTEVTTITGNGOPT-----DEVIIVIRTPPTAISSS 637
Qy 212 QSPMGONVKKLVEQLDAKIEAEFTRKLYVELKSSPOPLHVPF-----LKRSV- 262
Db 638 LSSSSGQITSI-----TSSRPIITPFPSNGSVISSVVIS 674
Qy 263 -ALRQLPNSQSFIOQCV-----OQTSMDVIACTTVTTSPVTTVSSSQ 309
Db 675 SSVTSISVTSFSSSVISSSTTTSTISFSESSTSVIPTSSSGSEKTSASASS 734
Qy 310 SEKSIIISGATAPRTVVOITNPLAGVGAAGVYTLHSVQPTAATGTTAGTGLQTSK 369
Db 735 SSSSISSESPKSP--TNSSSSLPV--TSATTGQETASSL--PPATTKTTSB---QTTL 784
Qy 370 PLVTSVANTVTVVSLQPEKPVVSGTAVTLLSPATFGETSGAALCLDSVKR----- 420
Db 785 VTVTSCSHCTESI--SSAIVSTAVTVS--GVTTETTCPTSTETTCOTGTTBOT 840
Qy 421 -----VSEFCMDHICKPVIGTPVOIKLAOPGVLSOPA---GIPF--- 457
Db 841 KGTTEOTTETTKQTVVTVTISCESDICS-----KTASPAIVSTSTATINCVTEYT 891
Qy 458 -----GSSSQQLFSLRHVVOQPSG--GNEKOVYTISSSTLTITQCGCKTMTPTIPTS 510
Db 892 TWCPSTSTESKOQTLVTVVTSCEGVCSETPSPAIIVSTATATVN-----DVTVYPTW 944
Qy 511 QPPASIIKQTLTPGNKILSLQASPTOKNRKEWNTS 547
Db 945 R-PQTMDSVSSKMSNSATSETTNTTGAETKTAVTS 980

RESULT 11
US-09-801-368-106
; Sequence 106, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 106
```

```
; LENGTH: 1169
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106

Query Match 6.28; Score 172; DB 10; Length 1169;
Best Local Similarity 21.48; Pred. No. 0.0011;
Matches 112; Conservative 85; Mismatches 191; Indels 136; Gaps 24;

Qy 4 VTKAVPAPPKVSSGRLPAPQIVAVKAPMTTIOF--PAULQLPQGVILKNSGPIML 62
Db 689 VTSSSVSTPITSSESSASVTLI---PSTISEKPSMK---TKVVISSSPTNL 740
Qy 63 VSPQOTVTRAETTSNITSRPAPV-----PANPQVKTCTPVNSSQOLIKKVAATPV 112
Db 741 ITSYDTNKSQDVTSSSTSVSLISSISLPSYSKASSQEDIFHSISVSSGQALTFESSIKV 800
Qy 113 KK-----LAQIGTVVTVTPK---PSSVQSAVPTSVVTVTPGK 149
Db 801 SSSSESHRTSPYTSSESGIKSSGVELESTSTSFSEHETSTASTVQISSQFVTPSSP 860
Qy 150 LNTVTTLKPSLSG---ASSPNSNEPNLKAENSAVOINLSPMLENVKCKKFLAMLK 205
Db 861 ISTVA---PKSTGLNSQTESTNNSKRETSSENSASV----- 893
Qy 206 IACSGOSPEMGONVKKLVEQLDAKIEAEF--FTR---KLYVELKSSPOPLHVPFLK 258
Db 894 MPSSSATPKTG-----KVISDETSSGFSRDTTVYRMTSETPTN---EQ 936
Qy 259 KSVVALROLPLNSQSFIOQCVQOTSSDMVIACTTV-----SPVTTVSS 307
Db 937 TLLITVSSCEENSCS-----NTVSSAVSTATTINGITTEVTCPLSATLETVSK 989
Qy 308 SOS--EKSTIVGATAPRTVVOITNPL--LAGPVGAAGVYTLHSVQPTAATGTTAGL 364
Db 990 LSEKTTLLITVTSCEGVCSETASPAIVSTATVNDVTVVTSWPOATNKLAIVSSD- 1048
Qy 365 LQTSKPLVTSVANTVTVVSLQPEKPVVSGTAVTLLSPATFGETSGAALCLDSVKR 424
Db 1049 IENSKSKSPFSEAEFTSISRNNFVP--TSGTISIEHT--TITSMNSENDVNS----- 1101
Qy 425 CMDHICKPVIGTPVOIKLAQPGVLSOPAGIPT---GSSSKOL 464
Db 1102 ASEAVSSKSVTNPVILSVSQ-----OPRGTPASSMIGSSTASL 1139

RESULT 12
US-10-161-510-10
; Sequence 10, Application US/10161510
; Publication No. US20020192695A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: PIBS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-074C
; CURRENT APPLICATION NUMBER: US/10/161,510
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1056
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-510-10

Query Match 5.88; Score 159.5; DB 9; Length 1056;
Best Local Similarity 21.58; Pred. No. 0.0064;
Matches 120; Conservative 56; Mismatches 210; Indels 171; Gaps 21;
```


Thu Feb 20 16:36:32 2003

us-09-763-909-2_copy_1_552.rpb

Page 10

Search completed: February 16, 2003, 22:02:06
Job time : 39.8745 secs

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
22

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 194.796 Seconds

(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2-COPY_1-552

Perfect score: 2758
Sequence: 1 GLVTRVAPVAPKVSQSGP.....ASPTOKRKIKENTSCFIDE 552Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2758	100.0	801	1	PCT-US01-08631-39827
2	2758	100.0	852	21	US-09-763-909-2
3	2678	97.1	865	1	PCT-US02-25829-25
4	2665	96.6	843	1	PCT-US02-29964-410
5	2235	81.0	685	27	US-60-243-468-1271
6	547	19.8	737	1	PCT-US94-01114-16

	7	547	19.8	1051	1	PCT-US02-30474-1660	Sequence 1660, Ap
	8	547	19.8	1051	27	US-60-324-631-1665	Sequence 1665, Ap
	9	535	19.4	1083	1	PCT-US02-30474-3271	Sequence 3271, Ap
	10	521	18.9	1023	1	PCT-US01-20592-14	Sequence 14, Ap
	11	521	18.9	1023	1	PCT-US01-20592A-14	Sequence 14, Ap
	12	521	18.9	1023	22	US-09-893-519A-14	Sequence 14, Ap
	13	258	9.4	899	27	US-60-167-217-11401	Sequence 11401, A
	14	258	9.4	921	27	US-60-173-464-9223	Sequence 9223, Ap
	15	258	9.4	921	27	US-60-173-464-9223	Sequence 9223, Ap
	16	256	9.3	921	1	PCT-US94-01114-2	Sequence 20611, A
	17	256	9.3	921	4	US-08-013-412-2	Sequence 2, Ap
	18	256	9.3	921	20	US-09-614-150-11376	Sequence 11376, A
	19	256	9.3	921	20	US-09-614-150-24957	Sequence 24957, A
	20	256	9.3	921	27	US-60-191-637-11408	Sequence 11408, A
	21	256	9.3	921	27	US-60-191-637-25058	Sequence 25058, A
	22	256	9.3	921	27	US-60-191-637-25058	Sequence 8933, Ap
	23	256	9.3	921	27	US-60-191-681-18933	Sequence 19739, A
	24	247.5	9.0	1953	27	US-60-185-361-566	Sequence 566, App
	25	247.5	9.0	2947	27	US-60-185-361-494	Sequence 494, App
	26	247.5	9.0	2947	27	US-60-185-361-938	Sequence 938, App
	27	237	8.6	1149	27	US-60-185-361-565	Sequence 565, App
	28	233.5	8.5	528	22	US-09-840-746-20	Sequence 20, App
	29	224	8.1	564	18	US-09-417-507-42234	Sequence 42234, A
	30	220.5	8.0	4742	26	US-10-221-279-12349	Sequence 12349, A
	31	220	8.0	1367	18	US-09-487-558-108	Sequence 108, App
	32	220	8.0	1367	22	US-09-487-558B-108	Sequence 108, App
	33	220	8.0	1367	22	US-09-801-368-108	Sequence 108, App
	34	220	8.0	1367	27	US-60-087-236-55	Sequence 55, App
	35	219	7.9	695	20	US-09-641-377-694	Sequence 694, App
	36	219	7.9	745	20	US-09-641-377-684	Sequence 684, App
	37	219	7.9	745	20	US-09-641-377-696	Sequence 696, App
	38	219	7.9	5179	19	PCT-US02-17382-218	Sequence 218, App
	39	219	7.9	5179	19	US-09-538-092-1258	Sequence 1258, Ap
	40	219	7.9	5179	20	US-09-609-448-11068	Sequence 1068, Ap
	41	219	7.9	5179	20	US-09-649-811-1068	Sequence 1068, Ap
	42	219	7.9	5179	22	US-09-833-263-1068	Sequence 1068, Ap
	43	219	7.9	5179	23	US-09-922-217-1068	Sequence 1068, Ap
	44	219	7.9	5179	24	US-10-025-380-1068	Sequence 1068, Ap
	45	216	7.8	778	20	US-09-641-377-685	Sequence 685, App

ALIGNMENTS

RESULT 1
PCT-US01-08631-39827
Sequence 39827, Application PC/TUS0108631
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-049
CURRENT APPLICATION NUMBER: PCT/US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO: 39827
LENGTH: 801
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-08631-39827

Query Match 100.0%; Score 2758; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 2; 2e-215; Indels 0; Gaps 0;
Matches 552; Conservative 0; Mismatches 20;
OY 1 GLVTRVAPVAPKVSQSGPRLPAQIVAVKAPNTTIOFPANLQDPGTVLKSNSGPL 60
DB 1 GLVTRVAPVAPKVSQSGPRLPAQIVAVKAPNTTIOFPANLQDPGTVLKSNSGPL 60

```

QY      61  MIVSQQQVETRAEETISNITSRAPVAPNPQTVAIKCVPNSSSOLIKKVAATPVKKLAQIGT  120
Db      61  MIVSQQVETRAEETISNITSRAPVAPNPQTVAICTVPNSSSOLIKKVAATPVKKLAQIGT  120
QY      121  TVVTVTPKPRSSVQSAVAVPVTSVVTPBGPPLNTVTTLKPSLSGASSTPSNEPNKAENSA  180
Db      121  TVVTVTPKPRSSVQSAVAVPVTSVVTPBGPPLNTVTTLKPSLSGASSTPSNEPNKAENSA  180
QY      181  VOINTSPMLLENVKKCKNPLAMLILKACSGSPENGGVKKLVEQLDLPAKTEAEFEETR  240
Db      181  VOINTSPMLLENVKKCKNPLAMLILKACSGSPENGGVKKLVEQLDLPAKTEAEFEETR  240
QY      241  LYVELKSSDPHLVPLPKKSVVALNQLLBNSSGSEFIOQCVOQTSDDVNIATCTTVTVTSBV  300
Db      241  LYVELKSSDPHLVPLPKKSVVALNQLLBNSSGSEFIOQCVOQTSDDVNIATCTTVTVTSBV  300
QY      301  VTTTSSSSGSEKSIIVSGTAPRTSVYONLNLNLAGVYGAACVNTLHSGPTAAATCGTNA  360
Db      301  VTTTSSSSGSEKSIIVSGTAPRTSVYONLNLNLAGVYGAACVNTLHSGPTAAATCGTNA  360
QY      361  GTGLLOTSKRPLTVSVANTVTVTSLOPEKPVSGTAVTSLSPAVTGETSGAAILCLPSVPR  420
Db      361  GTGLLOTSKRPLTVSVANTVTVTSLOPEKPVSGTAVTSLSPAVTGETSGAAILCLPSVPR  420
QY      421  VVSFQMDHICRKVIGTTPVQIKLAQGPVLSQAGIPTGSSSQQLPSLFHVYQOOPSQGNK  480
Db      421  VVSFQMDHICRKVIGTTPVQIKLAQGPVLSQAGIPTGSSSQQLPSLFHVYQOOPSQGNK  480
QY      481  QVTTTSHSSTLTIQCGOKTMEPNTLIPTSGPPASILKQILPLGKILSLQASPTQKNR  540
Db      481  QVTTTSHSSTLTIQCGOKTMEPNTLIPTSGPPASILKQILPLGKILSLQASPTQKNR  540
QY      541  IKENTVSCFRDE 552
Db      541  IKENTVSCFRDE 552

RESULT 2
US-09-763-909-2
; Sequence 2, Application US/09763909
; GENERAL INFORMATION:
; APPLICANT: Dikstein, Rivaia
; APPLICANT: Yamlet-Hezi, Ayala
; TITLE OF INVENTION: A TRANSCRIPTION FACTOR TFIID SUBUNIT,
; TITLE OF INVENTION: TAF1105, POLYPEPTIDES, DNA ENCODING THEREFOR AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 13005/002001
; CURRENT APPLICATION NUMBER: US/09/763,909
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 852
; TYPE: PRN
; ORGANISM: Homo sapiens
; US-09-763-909-2

```

;; TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
;; FILE REFERENCE: PF-1146 PCT
;; CURRENT APPLICATION NUMBER: PCT/US02/25829
;; CURRENT FILING DATE: 2002-08-14
;; PRIOR APPLICATION NUMBER: US 60/313,111
;; PRIOR FILING DATE: 2001-08-17
;; PRIOR APPLICATION NUMBER: US 60/314,682
;; PRIOR FILING DATE: 2001-08-24
;; PRIOR APPLICATION NUMBER: US 60/314,756
;; PRIOR FILING DATE: 2001-08-24
;; PRIOR APPLICATION NUMBER: US 60/315,105
;; PRIOR FILING DATE: 2001-08-27
;; PRIOR APPLICATION NUMBER: US 60/316,751
;; PRIOR FILING DATE: 2001-08-31
;; PRIOR APPLICATION NUMBER: US 60/316,856
;; PRIOR FILING DATE: 2001-08-31
;; PRIOR APPLICATION NUMBER: US 60/328,185
;; PRIOR FILING DATE: 2001-10-05
;; NUMBER OF SEQ ID NOS: 66
;; SOFTWARE: PERL Program
;; SEQ ID NO 25
;; LENGTH: 865
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc.feature
;; OTHER INFORMATION: Incyte ID No: 4398735CD1
PCT-US02-25829-25

Query Match 97.1%; Score 2678; DB 1; Length 865;
Best Local Similarity 97.7%; Pred. No. 8.3e-209;
Matches 542; Conservative 0; Mismatches 5; Indels 8; Gaps 2;

OY 1 GTLVTKVAPVAPKPVSSGPRLPAPQIVAVKAPMTTIIQFPANLQLP---GTVLIRKNS 57
DB 67 GTLVTKVAPVAPKPVSSGPRLPAPQIVAVKAPMTTIIQFPANLQLP---GTVLIRKNS 126
OY 58 GPLMLVSPQOVTBRAETTSNTSRPAVPANQVYKICTVPSNSQLIKKVAVTPVKKLAQ 117
DB 127 GPLMLVSPQOVTBRAETTSNTSRPAVPANQVYKICTVPSNSQLIKKVAVTPVKKLAQ 186
OY 118 IGTVVTVTKRPSVQAVAPTSTVYTPGKPLNTVTTLKPSIGASTPSNEPNLKAEN 177
DB 187 IGTVVTVTKRPSVQAVAPTSTVYTPGKPLNTVTTLKPSIGASTPSNEPNLKAEN 246
OY 178 SAAVOIMLSPTMLENVKCKNFMLMLIKLACSGSOSPEMGONVKKLVEQLDAKIEAEF 237
DB 247 SAAVOIMLSPTMLENVKCKNFMLMLIKLACSGSOSPEMGONVKKLVEQLDAKIEAEF 306
OY 238 TRKLYELKSSPOHVLVPLFKSVVALROLPLNSOSFIQOCVQOTSSDMVATCTTVT 297
DB 307 TRKLYELKSSPOHVLVPLFKSVVALROLPLNSOSFIQOCVQOTSSDMVATCTTVT 366
OY 298 SPVYTTVVSSQSKSTIISGATAPRIVSVQTLNPLAGPVGAKGVYTLHSVGFPAATGG 357
DB 367 SPVYTTVVSSQSKSTIISGATAPRIVSVQTLNPLAGPVGAKGVYTLHSVGFPAATGG 426
OY 358 TTACTGTLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLTSLAVYFGEESGAICLPS 417
DB 427 TTACTGTLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLTSLAVYFGEESGAICLPS 486
OY 418 VKPVVSCWMDHICKPVIGTIPVQIKLAOPGPVLSQPAIGTSSSKOLFSLFHVHQPSGG 477
DB 487 VKPVVSCWMDHICKPVIGTIPVQIKLAOPGPVLSQPAIGTSSSKOLFSLFHVHQPSGG 541
OY 478 NEKVYTISSSTLTICGCGKTPMVTIIPTSQFPASIIKQITLPGNKILSLQASPTQ 537
DB 542 NEKVYTISSSTLTICGCGKTPMVTIIPTSQFPASIIKQITLPGNKILSLQASPTQ 601
OY 538 KNRIKENVTSQCFRDE 552
DB 602 KNRIKENVTSQCFRDE 616

RESULT 4
PCT-US02-29964-410
Sequence 410, Application PC/TUS0229964
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Ren, Felyan
APPLICANT: Zhang, Jie
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Xue, Aiding J.
APPLICANT: Wang, Dunrui
APPLICANT: Ghosh, Malabika
APPLICANT: Asundi, Vinod
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Goodrich, Ryle W.
APPLICANT: Weng, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 809ACIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/29964
CURRENT FILING DATE: 2002-09-19
PRIOR APPLICATION NUMBER: US 60/323,739
PRIOR FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 922
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO 410
LENGTH: 843
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-29964-410
Query Match 96.6%; Score 2665; DB 1; Length 843;
Best Local Similarity 97.8%; Pred. No. 9.2e-208;
Matches 540; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 1 GTLVTKVAPVAPKPVSSGPRLPAPQIVAVKAPMTTIIQFPANLQLP---GTVLIRKNSGPT 60
DB 43 GTLVTKVAPVAPKPVSSGPRLPAPQIVAVKAPMTTIIQFPANLQLP---GTVLIRKNSGPT 102
OY 61 MLVSPQOVTBRAETTSNTSRPAVPANQVYKICTVPSNSQLIKKVAVTPVKKLAQIGT 120
DB 103 MLVSPQOVTBRAETTSNTSRPAVPANQVYKICTVPSNSQLIKKVAVTPVKKLAQIGT 162
OY 121 TVVTVTKRPSVQAVAPTSTVYTPGKPLNTVTTLKPSIGASTPSNEPNLKAENSA 180
DB 163 TVVTVTKRPSVQAVAPTSTVYTPGKPLNTVTTLKPSIGASTPSNEPNLKAENSA 222
OY 181 VOIMLSPTMLENVKCKNFMLMLIKLACSGSOSPEMGONVKKLVEQLDAKIEAEFTRK 240
DB 223 VOIMLSPTMLENVKCKNFMLMLIKLACSGSOSPEMGONVKKLVEQLDAKIEAEFTRK 282

```

QY 241 LVEELKSSPQPHLVPLFLKSSVVALRQLLPNSOSFIOQCVOQTSDDWVIACTTTVTSPV 300
DB 283 LVEELKSSPQPHLVPLFLKSSVVALRQLLPNSOSFIOQCVOQTSDDWVIACTTTVTSPV 342
QY 301 VTTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVVTLSHSGPTAATGTTA 360
DB 343 VTTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVVTLSHSGPTAATGTTA 402
QY 361 GTGLLOTSKPLVTSVANTVTVTSLOPEKPVSGTAATLSLPAVTFETSGAALCLPSVKR 420
DB 403 GTGLLOTSKPLVTSVANTVTVTSLOPEKPVSGTAATLSLPAVTFETSGAALCLPSVKR 462
QY 421 VVSFCDHICKPIGTPVQIKLAQPGPVLSQPGAGIPTGSSSKOLFSLFHVQOPSGGNEK 480
DB 463 VVSAGCTSDKPIGTPVQIKLAQPGPVLSQPGAGIPQAVQKOLFSLFHVQOPSGGNEK 522
QY 481 QVTTISHSSTLTIOKCGOKTMPVNTIIPTSQFPASILKQITLPGKKIISLOASPTQOKR 540
DB 523 QVTTISHSSTLTIOKCGOKTMPVNTIIPTSQFPASILKQITLPGKKIISLOASPTQOKR 582
QY 541 IKENVTSCFRDE 552
DB 583 IKENVTSCFRDE 594

```

RESULT 5
US-60-243-468-1271
; Sequence 1271, Application US/60243468

```

; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,
; FILE REFERENCE: C1000929
; CURRENT APPLICATION NUMBER: US/60/243,468
; CURRENT FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2121
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1271
; LENGTH: 685
; TYPE: PRP
; ORGANISM: HUMAN
US-60-243-468-1271

```

Query Match 81.0%; Score 2235; DB 27; Length 685;
Best Local Similarity 86.4%; Pred. No. 8,5e-173;
Matches 465; Conservative 6; Mismatches 25; Indels 42; Gaps 3;

```

QY 1 GTLVTVAVASAPRKVSSGPRLPAPQIVAVKANNTTIOFPANLOLPRTGLVLIKSSGRL 60
DB 67 GTLVTVAVASAPRKVSSGPRLPAPQIVAVKANNTTIOFPANLOLPRTGLVLIKSSGRL 126
QY 61 MLVSPQQTVTRABETSNITSRPAVPANPQTVKICTVPPNSSQLIKKVAATPVKKLAQIGT 120
DB 127 MLVSPQQTVTRABETSNITSRPAVPANPQTVKICTVPPNSSQLIKKVAATPVKKLAQIGT 163
QY 121 TVVTVTPKPPSSVQSAVAVPTSVTVTPGKPLNTVTLKPSLSGASSPSPNEPNKAENSA 180
DB 164 -----SVAVPTSVTVTPGKPLNTVTLKPSLSGASSPSPNEPNKAENSA 210
QY 181 VOINISPTMLENKKCNKFLAMLIKACGSSOSPENGQWKKIYEQDLDAKTAEFETR 240
DB 211 VOINISPTMLENKKCNKFLAMLIKACGSSOSPENGQWKKIYEQDLDAKTAEFETR 270
QY 241 LVEELKSSPQPHLVPLFLKSSVVALRQLLPNSOSFIOQCVOQTSDDWVIACTTTVTSPV 300
DB 271 LVEELKSSPQPHLVPLFLKSSVVALRQLLPNSOSFIOQCVOQTSDDWVIACTTTVTSPV 330
QY 301 VTTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVVTLSHSGPTAATGTTA 360
DB 331 VTTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVVTLSHSGPTAATGTTA 390

```

```

QY 361 GTGLLOTSKPLVTSVANTVTVTSLOPEKPVSGTAATLSLPAVTFETSGAALCLPSVKR 420
DB 391 GTGLLOTSKPLVTSVANTVTVTSLOPEKPVSGTAATLSLPAVTFETSGAALCLPSVKR 450
QY 421 VVSFCDHICKPIGTPVQIKLAQPGPVLSQPGAGIPTGSSSKOLFSLFHVQOPSGGNEK 480
DB 451 VVSAGCTSDKPIGTPVQIKLAQPGPVLSQPGAGIPQAVQKOLFSLFHVQOPSGGNEK 505
QY 481 QVTTISHSSTLTIOKCGOKTMPVNTIIPTSQFPASILKQITLPGKKIISLOASPTQOKR 538
DB 506 QVTTISHSSTLTIOKCGOKTMPVNTIIPTSQFPAGK-KHIDITLNSDAVLISQATQOE 562

```

RESULT 6
PCT-US94-01114-16
; Sequence 16, Application PC/TUS9401114

```

; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

```

```

; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01114
; FILING DATE: 28-JAN-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-57650-2/AUT/RMO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01114-16

```

Query Match 19.8%; Score 547; DB 1; Length 737;
Best Local Similarity 30.2%; Pred. No. 6,4e-35;
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21;

```

QY 13 PPKVSSG-----RLPAPQIVAVKANNTTIOFPANLOLPRTGLVLIKSSGRL 61
DB 68 PRTATSGIATLTPYLARLRP-----PQPTNIO---NQQLPQAVLVASEGQL 118
QY 62 LVSPQQTVTR-----AETSNITSRPAVPANPQTVKICTVPPNSSQLIKKVAATPVKKL 115
DB 119 MI-PQALQMQAQAHAQPOTTMAPRPTSPAPVQISTVQAGPPIAR-QVTP----- 172
QY 116 AQIGTIVTVTPKPPSSVQSAVAVPTSVTVTPGKPLNTVTLKPSLSGASS-----TPS 168

```

```

Db 173 -----TTIKOV---SQAQTVPQATLQSRSPGVQPOLVLGGAQTALGTATAVGTGPQ 225
QY 169 NE-PNLAKENSAAVOINLSPMLTENVKCKNFMLIKLACSGSGSPGEMGNVKKLEVL 227
Db 226 RTVGATTTSSAATE-----TMENVKCKNFSTLTKLASSGSGSTETANVKEVLONL 279
QY 228 LDAKIEAEFTFKLYELKSSPOPHVLPFLKRSVVALROLPLNSQSFIOOCVQO-----TS 283
Db 280 LDGKIEADEFTSRLYRELNSPPQYLVPLFKRSLPALROLTPDAAFIQSQOQPPPTPS 339
QY 284 SDVIAATCTTIVTSPVTTTSSQSEKSLIYSGATAPRTVSQTLNPLAGPVGAKAGV 343
Db 340 Q-----ATTALTAVALSSSVORTAGTATATVTSALQPPVLSL----- 376
QY 344 VTLHSGVPTAAGTGTAGTGLQTSKPLVTSVANTVTVLSQPEKPVVSGTAVTSLSPAY 403
Db 377 -----TQPTQVGVKGQGPPLVIO-----QPKP-----GALIRPPQV 410
QY 404 TFGTSGAATCLPSVKPVSECFMDHICKPVYIGTPVQIKLAPGVLSOPAGIPGSSSKQ 463
Db 411 TLQGT-----PMVALROPH-NRIMLTTPQOIOI----- 437
QY 464 LFSLFHVQOPSGGENEKQVTTISHSSTLTIOKCGQKTMPTVTTIPTSQFPASTLKQITL 523
Db 438 ----- 552
QY 524 PGKRL---SLQASPTQKNRIKENVTSCFRDE 552
Db 452 PGTALSAVSNQAAAAQKNKKEPGGGSFRDD 483

```

RESULT 7

```

PCT-US02-30474-1660
; Sequence 1660, Application PC/TUS0230474
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aldong J.
; APPLICANT: Zhao, Qiang A.
; APPLICANT: Wang, Jien-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrul
; APPLICANT: Ma, Yundqing
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: PCT/US02/30474
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/324,631
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3476
; SOFTWARE: PL_Fl_genes Version 6.0
; SEQ ID NO 1660
; LENGTH: 1051
; TYPE: PRT
; ORGANISM: Homo sapiens
; PCT-US02-30474-1660

```

```

Query Match          19.8% Score 547; DB 1; Length 1051;
Best Local Similarity 30.2% Pred. No. 1,1e-34;
Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21;

```

```

QY 13 PPKVSG-----PLPAPQIVAKAPNTTTTQFPANLQPLPGTVLIRKNSGGLM 61
Db 382 PTATSGIRATILPTVLAPRLPQ-----PQNPNTIQ-----NFQLPQGVYLVRENGQL 432
QY 62 LVSPQGVTR-----AETSNITSRPAVAPNPQTVKICVYIPNSSSOLIKKAVTPVKKL 115
Db 433 MI-PQALAQQAQAHQAPQTTMAPRATPTSPAPVQISTVQAPGTPIAR-QVTP----- 486
QY 116 AQIGTVVTVTPKPSVQSVAVPRTSVTVTPGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 487 -----TTIKOV---SQAQTVPQATLQSRSPGVQPOLVLGGAQTALGTATAVGTGPQ 539
QY 169 NE-PNLAKENSAAVOINLSPMLTENVKCKNFMLIKLACSGSGSPGEMGNVKKLEVL 227
Db 540 RTVGATTTSSAATE-----TMENVKCKNFSTLTKLASSGSGSTETANVKEVLONL 593
QY 228 LDAKIEAEFTFKLYELKSSPOPHVLPFLKRSVVALROLPLNSQSFIOOCVQO-----TS 283
Db 594 LDGKIEADEFTSRLYRELNSPPQYLVPLFKRSLPALROLTPDAAFIQSQOQPPPTPS 653
QY 284 SDVIAATCTTIVTSPVTTTSSQSEKSLIYSGATAPRTVSQTLNPLAGPVGAKAGV 343
Db 654 Q-----ATTALTAVALSSSVQRTAGTATATVTSALQPPVLSL----- 690
QY 344 VTLHSGVPTAAGTGTAGTGLQTSKPLVTSVANTVTVLSQPEKPVVSGTAVTSLSPAY 403
Db 691 -----TQPTQVGVKGQGPPLVIO-----QPKP-----GALIRPPQV 724
QY 404 TFGTSGAATCLPSVKPVSECFMDHICKPVYIGTPVQIKLAPGVLSOPAGIPGSSSKQ 463
Db 725 TLQGT-----PMVALROPH-NRIMLTTPQOIOI----- 751
QY 464 LFSLFHVQOPSGGENEKQVTTISHSSTLTIOKCGQKTMPTVTTIPTSQFPASTLKQITL 523
Db 752 ----- 765
QY 524 PGKRL---SLQASPTQKNRIKENVTSCFRDE 552
Db 766 PGTALSAVSNQAAAAQKNKKEPGGGSFRDD 797

```

RESULT 8

```

; Sequence 1665, Application US/60324631
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aldong J.
; APPLICANT: Zhao, Qiang A.
; APPLICANT: Wang, Jien-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrul
; APPLICANT: Ma, Yundqing

```

```

APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhiwei
APPLICANT: Meng, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Dimañac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: 810
CURRENT APPLICATION NUMBER: US/60/324,631
CURRENT FILING DATE: 2001-09-24
PRIORITY APPLICATION NUMBER: US 09/488,725
PRIORITY FILING DATE: 2000-01-21
PRIORITY APPLICATION NUMBER: US 09/552,317
PRIORITY FILING DATE: 2000-04-25
PRIORITY APPLICATION NUMBER: PCT/US00/35017
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: US 09/491,404
PRIORITY FILING DATE: 2000-01-25
PRIORITY APPLICATION NUMBER: PCT/US01/02623
PRIORITY FILING DATE: 2001-01-25
PRIORITY APPLICATION NUMBER: US 09/496,914
PRIORITY FILING DATE: 2000-02-03
PRIORITY APPLICATION NUMBER: US 09/560,875
PRIORITY FILING DATE: 2000-04-27
PRIORITY APPLICATION NUMBER: PCT/US01/03800
PRIORITY FILING DATE: 2001-02-05
PRIORITY APPLICATION NUMBER: US 09/515,126
PRIORITY FILING DATE: 2000-02-28
PRIORITY APPLICATION NUMBER: US 09/577,409
PRIORITY FILING DATE: 2000-05-18
PRIORITY APPLICATION NUMBER: PCT/US01/04927
PRIORITY FILING DATE: 2001-02-26
PRIORITY APPLICATION NUMBER: US 09/519,705
PRIORITY FILING DATE: 2000-03-07
PRIORITY APPLICATION NUMBER: US 09/574,454
PRIORITY FILING DATE: 2000-05-19
PRIORITY APPLICATION NUMBER: PCT/US01/04941
PRIORITY FILING DATE: 2001-03-05
PRIORITY APPLICATION NUMBER: US 09/540,217
PRIORITY FILING DATE: 2000-03-31
PRIORITY APPLICATION NUMBER: US 09/649,167
PRIORITY FILING DATE: 2000-08-23
PRIORITY APPLICATION NUMBER: PCT/US01/08631
PRIORITY FILING DATE: 2001-03-30
PRIORITY APPLICATION NUMBER: US 09/552,929
PRIORITY FILING DATE: 2000-04-18
PRIORITY APPLICATION NUMBER: US 09/770,160
PRIORITY FILING DATE: 2001-01-26
PRIORITY APPLICATION NUMBER: PCT/US01/08656
PRIORITY FILING DATE: 2001-04-18
PRIORITY APPLICATION NUMBER: US 09/577,408
PRIORITY FILING DATE: 2000-05-18
PRIORITY APPLICATION NUMBER: PCT/US01/14827
PRIORITY FILING DATE: 2001-05-16
NUMBER OF SEQ ID NOS: 3334
SOFTWARE: pc_flt_genes Version 6.0
SEQ ID NO 1665
LENGTH: 1051
TYPE: PRT
ORGANISM: Homo sapiens
US-60-324-631-1665

19.8%; Score 547; DB 27; Length 1051;
Beat Local Similarity 30.2%; Pred. No. 1,1e-34;
Matches 173; Conservative 67; Mismatches 144; Indels 108; Gaps 21

13 PPKVSSG-----PRLPAQIVAVKADNNTTIOFPANLQLPCTGVILKSNSGPLM 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 382 PSTATSGIRATLPTVLARLPQP-----PONPTNIO--NFQLPGRVLRSENGQL 432

62 LVSPOQYTR-----AETSNITSRPAVPANPQYVTKCTVNSSQLTKKAAVTPVKL 115
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
433 MI-FOQALQMOQAQNPQTTMAVRPAVPSAPRVQISTVQAPGPIIAR-QVTP--- 486

```

```

0Y 116 A0IGTAVTTVPKPSQVQSAVNPVSVTVVPGKXNLNVT--TLKPSLIGASS-----TPS 168
Db 487 -----TTIKQV-----SQAOITVPQSTLORSBGVOPOLVLGGAOTASLGTATAVGTGPQ 339
0Y 169 NE-PNLKENSAAVOINLSPMLNEVKKCKNFKMLILKILKACSGSQSPDMGONVKKLVEOL 227
Db 540 RVPQATVTTSSAATE-----TMEVKKCKNFKLSTLILKILKASSGKQSTETANVKKELVQNL 593
0Y 228 IDAKLEAEFFRKLYLVELKSPQCHLYPLFLKSVYALROLLPNSQSTQOCVQO-----TS 283
Db 594 LDGKLEADFTFSRLYRELNSSPQYLVPLFKRLSLPALROLPLPSAAFLQOS000PREPTS 653
0Y 284 SDMVATCTTWTVTSPVTVTVSSSQSKSIIVSGATAPRTVSQTLNPLAPGVAKAGV 343
Db 654 Q-----ATTALRALRVLVSSSVQRTAGKTATATVTSALQPPVLSL-----600
0Y 344 VTLHSVGPPTAATGGTTAGTGLLOITSKPLVTSVANTVTVTVSLQPERKPVSGTAVTLSLPAV 403
Db 691 -----TOPVQVGVKGQGPPTLVLD-----QPKRP-----GALIRPPQV 724
0Y 404 TFGESGAALCLPSVKPVVSCWCHICKPVYGTVPQIKLAPGCVLVSLPAGIFPFGSSSKQ 463
Db 725 TLTQF-----PVALRQPH-NKIMLTTPQOJOL-----751
0Y 464 LSLFHVQVQPSGNEKQVTTIHSSTLTTLQCGQKTMPVVTIIPVQFPFASILKQITL 523
Db 752 -----NPLQVPVVKPAVL 765
0Y 524 PGNKIL--SLQASPTOKNIKRENTVSCFND 552
Db 766 PGTKALSAVSAQAAAAQKNKLEKPGCGSFRD 797

RESULT 9
PCT-US02-30474-3271
Sequence 3271, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aidong J.
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Duntui
APPLICANT: Ma, Yungqing
APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhiwei
APPLICANT: Wang, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 810CIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/30474
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: US 60/324, 631
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03

```



```

RESULT 11
PCT-US01-20592A-14
Sequence 14, Application PC/TUS0120592A
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G548-US1
CURRENT APPLICATION NUMBER: PCT/US01/20592A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PR
ORGANISM: Homo sapiens
NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592A-14

Query Match      18.9%; Score 521; DB 1; Length 1023;
Best Local Similarity 31.7%; Pred. No. 1,4e-32;
Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;

QY 13 PRVSSG-----PRLPAPQVAVKAPVTITIOFPANQLPFGYVLINSNGPLM 61
DB 414 PIAATGIRATLTPTVAPLPLP-----PQNPNTNQ--NQLPBGWLVSENGQL 464
QY 62 LVSPQOTVTR-----AETTSNITSRPVAVPANPOTVKTICVPRSSSOLIKKVAVTVPKKL 115
DB 465 MI-PQOALQMOQAHAQOTTMARPARPTSPAPVOISVQAPGPIITAR-QVTP----- 518
QY 116 AOGITVTVTPKPVSSVQVAVPTSVVYTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 519 ----TTIKQV---SOAQTVQPSATLQSPGVQPOLVLGGAQOTSLGTAFAVQGTGPQ 571
QY 169 NE-PNLKAENSAVQINLSPTMLENKKNFLAMLIKILACSSQSPBEGQONKRLVEOL 227
DB 572 RTVPGATTTSSATE-----TMENVKCKKNFLSTLIKILASSQOSTETAANYKELVONL 625
QY 228 LPAKIAEFTKRLVYELKSSPOPHLVPLFKKSVALROLPLPSOSFIOQVOQ-----TS 283
DB 626 LQKIEADEFTSLRYELKSSPOPHLVPLFKKSVALROLPLPSOSFIOQVOQ-----TS 283
QY 284 SDNVATACCTVTVTPKPVSSVQVAVPTSVVYTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 686 Q-----ATTALTAVVLSSSVQRTAGKTAATVTSALQPPVLST----- 722
QY 344 VTLHSGPRAAAGGTAGAGGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPAPV 403
DB 723 -----TQPTQVGVGKQGOQPPVLVQ-----OPPKP-----GALLIRPPV 756
QY 404 TREGTSGAALCLPSVKPVPVSWFCMDHICKPVGTPVOIKLAQPGPVL-----SOPAGIPTGS 459

```

```

DB 757 TLTQT-----PMVALROPH-NRIMLTTPQOVNLSSESRILATNSLVGLTIRS 804
QY 460 SSQQLSELFHVQ-----OPSGCNEKQ---VTTISHSSTLTATQKQOR 499
DB 805 CKDETFLQAPLQRIELICKKHGITEHLPDVSVSHATQORLQNLVER 854

RESULT 12
US-09-893-519A-14
Sequence 14, Application US/09893519A
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PR
ORGANISM: Homo sapiens
NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CA472189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)
US-09-893-519A-14

Query Match      18.9%; Score 521; DB 22; Length 1023;
Best Local Similarity 31.7%; Pred. No. 1,4e-32;
Matches 168; Conservative 73; Mismatches 157; Indels 132; Gaps 22;

QY 13 PRVSSG-----PRLPAPQVAVKAPVTITIOFPANQLPFGYVLINSNGPLM 61
DB 414 PIAATGIRATLTPTVAPLPLP-----PQNPNTNQ--NQLPBGWLVSENGQL 464
QY 62 LVSPQOTVTR-----AETTSNITSRPVAVPANPOTVKTICVPRSSSOLIKKVAVTVPKKL 115
DB 465 MI-PQOALQMOQAHAQOTTMARPARPTSPAPVOISVQAPGPIITAR-QVTP----- 518
QY 116 AOGITVTVTPKPVSSVQVAVPTSVVYTPGKPLNTVT--TLKPSLGLASS-----TPS 168
DB 519 ----TTIKQV---SOAQTVQPSATLQSPGVQPOLVLGGAQOTSLGTAFAVQGTGPQ 571
QY 169 NE-PNLKAENSAVQINLSPTMLENKKNFLAMLIKILACSSQSPBEGQONKRLVEOL 227
DB 572 RTVPGATTTSSATE-----TMENVKCKKNFLSTLIKILASSQOSTETAANYKELVONL 625
QY 228 LPAKIAEFTKRLVYELKSSPOPHLVPLFKKSVALROLPLPSOSFIOQVOQ-----TS 283
DB 626 LQKIEADEFTSLRYELKSSPOPHLVPLFKKSVALROLPLPSOSFIOQVOQ-----TS 283

```

QY 284 SDVIAICTTTVTSPVTTTSSOSSEKSIIVSGATAAPRTVSQVTLNPLAGPVGAKGV 343
D 686 Q-----ATTALNAVVLSSSVORTAGKATATVLSALQPVLSL----- 722
QY 344 VTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTTSIQPEKPVVSGTAVTLSPAV 403
D 723 -----TQPTQVGVGOGQPTPLVIO-----QPPKP-----GALIRPQV 756
QY 404 TFGTSGAALCPVSKPVVSCFMDHICKPVIGTVPQIKLAQPGVPL-----SQPAGIPGTS 459
D 757 TLVQT-----PVALRQPH-NRIMLTTPQVVLSESRILATNSELVCTLRS 804
QY 460 SSKOLFSLFHVQ-----OPSGNEKQ-----VTTSHSSTLTQKCGQK 499
D 805 CKDEFLLQAPLQRRILEIGKKGITELHPDVVSYSVSHATQORLQNVLEK 854

RESULT 13
US-60-167-217-11401
Sequence 11401, Application US/60167217
GENERAL INFORMATION:
APPLICANT: Li, Peter W. D.
TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
FILE REFERENCE: CLO00152
CURRENT APPLICATION NUMBER: US/60/167, 217
CURRENT FILING DATE: 1999-11-24
NUMBER OF SEQ. ID NOS: 23195
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 11401
LENGTH: 899
TYPE: PRT
ORGANISM: Drosophila
US-60-167-217-11401

Query Match 9.4%; Score 258; DB 27; Length 899;
Best Local Similarity 22.9%; Pred. No. 3.5e-11;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 POIYAVKAPNTTIOFPANLQLPCTVILIKNSGPIML-----VSPQOTVRAETTS 76
D 111 POSBITITLNTGQTPA-----LVKTDNGFQLRVGTTGPTTGTITNTSNN 162
QY 77 NITSRPAVPANPQTVKICTVENSQ-----LIKKVAVTPVKKLAQIGTVVTVTP 127
D 163 NTSTTNHPTTQ-IRLQTVPAASMTNTATSNIIIVSVASSGYANSQPPHLLQMAQ 221
QY 128 KPSSVQSAVPTSVYTVTPGKPLNTVTTLKPSLSGASSTPSNEPMLKAENSAVOINLSP 187
D 222 APQLPQITQITIPAQSQOQOVNNVSSAGTAAVASTTA-----ATT 265
QY 188 TMLENVK-KCNFLAMLKILACSGSPGEMGNVKKLEQLLDKIEAEFTRKLYELK 246
D 266 TQGNTEKCKRFLANLEL-STREPKPEKVNVTLLQELVNAVVEEFCDRLERLN 323
QY 247 SSPQHLVPLFKSVVALROL-----LPSQSFIO----- 276
D 324 ASPQCLIGFLKKSPLRLQALYTKELYEGIKRPPQHVGLAGLSQQLPRIQAOIRPIG 383
QY 277 -----QCVQOTSSDMVIATCTTTVTSPVTTTSSOSEKSIIVSGATAPRTVS---VOT 329
D 384 PSQTTTIGQTOVRMI--TPNALGTPRPTIGHTTISKQPPN---IRLPAPRLVNTGRT 438
QY 330 LNPLAGPVGAKGVVTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTTSIQPEK 389
D 439 QIP-SLOVPGGANIVQIR--GPOHAQLORTGSVOIRATRP-----PNSVPTAN----- 484
QY 390 VVSGTAVTSLPVAFTGTSAGALCLPSVKPVVSCFMDHICKPVIGTVPQIKLAQPGVPL 449
D 485 -----KLNAVVGOTOIKAI--TPSLHP-----PSL 508

QY 450 SOPACIPGSSSKOLFSLFHVVOQPSGNEKQVTTTSHSSTLTQKCGQKTMPTVNTIIP 509
D 509 AAIISGPPPTLTVLS-----TLNSAST-----TLPIPS-LPT 542
QY 510 SOPPASI-----LKQITLPGN-----KILSLQA---SPTQKNRIKENVT 546
D 543 VHLPEALRAREQMONSLNHSNHPDAKLVEIKAPSLHPHMERINAST 592

RESULT 14
US-60-173-464-9223
Sequence 9223, Application US/60173464
GENERAL INFORMATION:
APPLICANT: Li, Peter W. D.
TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
FILE REFERENCE: CLO00173
CURRENT APPLICATION NUMBER: US/60/173, 464
CURRENT FILING DATE: 1999-12-29
NUMBER OF SEQ. ID NOS: 30269
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 9223
LENGTH: 921
TYPE: PRT
ORGANISM: Drosophila
US-60-173-464-9223

Query Match 9.4%; Score 258; DB 27; Length 921;
Best Local Similarity 22.9%; Pred. No. 3.6e-11;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 POIYAVKAPNTTIOFPANLQLPCTVILIKNSGPIML-----VSPQOTVRAETTS 76
D 133 POSBITITLNTGQTPA-----LVKTDNGFQLRVGTTGPTTGTITNTSNN 184
QY 77 NITSRPAVPANPQTVKICTVENSQ-----LIKKVAVTPVKKLAQIGTVVTVTP 127
D 185 NTSTTNHPTTQ-IRLQTVPAASMTNTATSNIIIVSVASSGYANSQPPHLLQMAQ 243
QY 128 KPSSVQSAVPTSVYTVTPGKPLNTVTTLKPSLSGASSTPSNEPMLKAENSAVOINLSP 187
D 244 APQLPQITQITIPAQSQOQOVNNVSSAGTAAVASTTA-----ATT 287
QY 188 TMLENVK-KCNFLAMLKILACSGSPGEMGNVKKLEQLLDKIEAEFTRKLYELK 246
D 288 TQGNTEKCKRFLANLEL-STREPKPEKVNVTLLQELVNAVVEEFCDRLERLN 345
QY 247 SSPQHLVPLFKSVVALROL-----LPSQSFIO----- 276
D 346 ASPQCLIGFLKKSPLRLQALYTKELYEGIKRPPQHVGLAGLSQQLPRIQAOIRPIG 405
QY 277 -----QCVQOTSSDMVIATCTTTVTSPVTTTSSOSEKSIIVSGATAPRTVS---VOT 329
D 406 PSQTTTIGQTOVRMI--TPNALGTPRPTIGHTTISKQPPN---IRLPAPRLVNTGRT 460
QY 330 LNPLAGPVGAKGVVTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTTSIQPEK 389
D 461 QIP-SLOVPGGANIVQIR--GPOHAQLORTGSVOIRATRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPVAFTGTSAGALCLPSVKPVVSCFMDHICKPVIGTVPQIKLAQPGVPL 449
D 507 -----KLNAVVGOTOIKAI--TPSLHP-----PSL 530
QY 450 SOPACIPGSSSKOLFSLFHVVOQPSGNEKQVTTTSHSSTLTQKCGQKTMPTVNTIIP 509
D 531 AAIISGPPPTLTVLS-----TLNSAST-----TLPIPS-LPT 564
QY 510 SOPPASI-----LKQITLPGN-----KILSLQA---SPTQKNRIKENVT 546
D 565 VHLPEALRAREQMONSLNHSNHPDAKLVEIKAPSLHPHMERINAST 614

```
RESULT 15
; Sequence 20611, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20611
; LENGTH: 921
; TYPE: PRF
; ORGANISM: Drosophila
; US-60-173-464-20611

Query Match          9.4%; Score 258; DB 27; Length 921;
Best Local Similarity 22.9%; Pred. No. 3,6e-11;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 POVAKAPNTTIOFPANQLPBGVLIKSNGLML-----VSPQOTVRAETTS 76
DB 133 POSPTITLSTLNTGQTPA-----LVKTDNGFOLRVGTTGPTTYQTITNTSNNS 184
QY 77 NITSRAVPANPQTVKICVNPSSQ-----LIKVAVTPVKLAQIGTVVTVTP 127
DB 185 NTSSTNHPTTQ-IRLQVPAASMTNTATSNITIVNSVASSGVANSSQPHLTQLMAQ 243
QY 128 KPSVSVAVPFSVTVTPEKPLNTYTTIKPSSLAGSSTPSNEPNLKAENSAVQINLSP 187
DB 244 APQLPQITQITIPAOQSOQOVNNSAGCTFAVSSSTA-----ATT 287
QY 188 TMLDNK-KCNFLAMLKILACSGSOPSEMGQVKKIVBQLDAKTEAEETKLYVELK 246
DB 288 TOGNTKCKKRLANLIEL--STREPKVEKAVRLLIQELVANVPEEFCRLERLNL 345
QY 247 SSBQPHLVPELKKSVVALRQL-----LPNSQSFQ--- 276
DB 346 ASPQCLIGFLKSLPLRLQALYTKELVIEGKPPQHVLAGLAGLSQQLPKIQAIPIG 405
QY 277 ----QCVQOTSSDMVATCTTYYTISPVVTVTVSSSQSEKSIIVSGATAPRTVS---VOT 329
DB 406 PSQTTIGQTVQVMI--TPNALGTTPRTIGHTTISKOPN--IRLPTARLVNTGIGRT 460
QY 330 LNPFLAGVGAAGVTVLHNSVGPATAGTGTAGGLQTSKPLVTSVANTVTVTSLOPEK 389
DB 461 QTP-SLQVPGQANIVQIR--GRQHAQLOKRTGSQVIRATIRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLSPAVTGETSGAAILPSVKPVVSVFCWDHICKPVGIGTPVOIKLAQPGVL 449
DB 507 -----KLTAVKVGOIQIKAI-TPSLHP-----PSL 530
QY 450 SOPAGITPTGSSSKQLSLRHVVOQPSGNGEKQVTTISHSSTLTIOKCGAKTMPTVNTIIP 509
DB 531 AAISSGPPPTPLSLVS-----TLNSAST-----TTLPIPS-LPT 564
QY 510 SGPFPASI-----LKQITLPGN-----KILSLQA---SPTOKRIKENVT 546
DB 565 VHLPPALAREQMOMSLNHNHNPDAKLVEIKAPSLRPHMERINASLT 614
```

Search completed: February 16, 2003, 22:18:18
Job time : 204.796 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:00:19 ; Search time 22.4582 Seconds
(without alignments)
2076.603 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552
Perfect score: 2758
Sequence: 1 GILVTRVAPVSAADPKVSSGP.....ASPTQNRKIKENTSCPRDE 552

Scoring table:
BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 429898 seqs, 84487048 residues

Total number of hits satisfying chosen parameters: 429898

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	539	19.5	709	US-09-724-676-63927	Sequence 63927, A
2	539	19.5	709	US-09-724-676-63927	Sequence 63927, A
3	539	19.5	741	US-09-724-676-63930	Sequence 63930, A
4	539	19.5	805	US-09-724-676-63921	Sequence 63921, A
5	539	19.5	805	US-09-724-676-63921	Sequence 63921, A
6	539	19.5	837	US-09-724-676-63924	Sequence 63924, A
7	539	19.5	837	US-09-724-676-63924	Sequence 63924, A
8	539	19.5	925	US-09-724-676-63928	Sequence 63928, A
9	527	19.1	925	US-09-724-676-63928	Sequence 63928, A
10	527	19.1	957	US-09-724-676-63931	Sequence 63931, A
11	527	19.1	1021	US-09-724-676-63922	Sequence 63922, A
12	527	19.1	1021	US-09-724-676-63922	Sequence 63922, A
13	527	19.1	1053	US-09-724-676-63925	Sequence 63925, A
14	527	19.1	1053	US-09-724-676-63925	Sequence 63925, A
15	527	19.1	679	US-09-724-676-63929	Sequence 63929, A
16	527	19.1	679	US-09-724-676-63929	Sequence 63929, A
17	519	18.8	711	US-09-724-676-63920	Sequence 63920, A
18	519	18.8	711	US-09-724-676-63920	Sequence 63920, A
19	519	18.8	775	US-09-724-676-63923	Sequence 63923, A
20	519	18.8	775	US-09-724-676-63926	Sequence 63926, A
21	519	18.8	807	US-09-724-676-63926	Sequence 63926, A
22	519	18.8	807	US-09-724-676-63926	Sequence 63926, A
23	519	18.8	5179	US-60-438-735-151	Sequence 151, App
24	519	18.8	5179	US-60-438-735-151	Sequence 151, App
25	206.5	7.5	1322	US-10-218-140-4828	Sequence 4828, Ap
26	206.5	7.5	1322	US-10-218-140-4828	Sequence 4828, Ap

27	205.5	7.5	2108	1	PCT-US02-33723-2	Sequence 2, Appli
28	205.5	7.5	2135	6	US-10-288-798-9	Sequence 9, Appli
29	205.5	7.5	2382	6	US-10-196-935A-2	Sequence 2, Appli
30	205.5	7.5	2382	6	US-10-293-017-48	Sequence 48, Appli
31	205.5	7.5	2382	6	US-10-052-648A-40	Sequence 40, Appli
32	205.5	7.5	2382	6	US-10-293-071-48	Sequence 48, Appli
33	201.5	7.3	5374	6	US-10-028-248A-75	Sequence 75, Appli
34	201.5	7.3	5374	6	US-10-107-782-75	Sequence 75, Appli
35	199.5	7.2	2601	6	US-10-028-248A-76	Sequence 76, Appli
36	199.5	7.2	2601	6	US-10-107-782-76	Sequence 76, Appli
37	197.5	7.2	5376	6	US-10-028-248A-74	Sequence 74, Appli
38	197.5	7.2	5376	6	US-10-107-782-74	Sequence 74, Appli
39	194.5	7.1	2781	6	US-10-263-929-122	Sequence 122, App
40	194	7.0	2971	6	US-10-218-140-1990	Sequence 1990, Ap
41	192.5	7.0	1794	7	US-60-427-045-299	Sequence 299, App
42	192.5	7.0	1799	7	US-60-427-045-149	Sequence 149, App
43	192.5	7.0	1821	7	US-60-427-045-82	Sequence 82, Appli
44	192.5	7.0	11721	7	US-60-427-045-162	Sequence 162, App
45	190.5	6.9	5935	6	US-10-243-243A-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-724-676-63927
: Sequence 63927, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: CompuGen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 CompuGen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 63927
: LENGTH: 709
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676-63927

Query Match 19.5%, Score 539, DB 5, Length 709;
Best Local Similarity 30.2%, Pred. No. 3.1e-22;
Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

QY	13	PKVSSG-----PRLPAPQIVAKA-----NTTTPRALQOLPPGTVLIKNSGRLM 61
DB	286	PTATTSGLRATLPTVLAIRLPQ-----PONPTNIO---NFQLPQGVILVRSNGQL 336
QY	62	LVSPOQVTR-----AETTSNITSRPAPVAPQVFKICTVPSNSSOLIKKVAVPPVKL 115
DB	337	MI-PQALAQMAQMAHPQPTMAPRAPPTPSAPVQISTVQAPPTPIAR-QVTP----- 390
QY	116	AOIGTTVTVTPKSSVSVAVPTISVTVTPCKPLNTVT--TLKPSLGASS-----TPS 168
DB	391	---TTIIKQV---SQAOTVQPSATLQRPQVQQLVIGMAAQTAIGTATAVQGTPIQ 443
QY	169	NE-PLKENSNAVQINISPTLLENVKKCKNTLAILILACSGSSPEMGQVAKVLVEOL 227
DB	444	RTVPATTTSSAAT-----TMENVKCKNKLSTLIKASSGKOSTFAANVKELVONTL 497
QY	228	IDAKTEAEETPRKLYVELKSPQHLVPLFKSVAVALPNSOSFIQOCVQO-----TS 283
DB	498	LDGKTEADEPFSRLYRELNSSQPLVLPFKLSLALQOLPDSAFITQOSQOQPPPTPS 557
QY	284	SDMVIATCTTTVTSSVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKGV 343
DB	558	Q-----ATTALPFAVLVSSSVQRTAGKTAATVTSALQPVLSL----- 594
QY	344	VTLSHVGFPATAGTGTAGTGLQTSKPLVTSANVTTVSLQPEKPVVSGRAVITSLAV 403
DB	595	-----TQPVQGVGKQGOPTPLVIO-----QPKP-----GALIRPQV 628

```
OY 404 TFGTSGAALCLPSVKPVSEFCMDHICKPVIGTPVOIKLAQPGPVLSQAPAGIPTGSSSKQ 463
DB 629 TLITQT-----PMVALRQPH-NRIMLTTPQOIQ----- 655
OY 464 LFSLFHVVOQPSGSGNEKOVYTTISHSSTLTIOKCGOKTMTPTIIPISQFPASILKOITL 523
DB 656 -----NPLQPVVVKRAVL 669
OY 524 PGKIL---SLQASPTQKNRIKENVTSCFR 550
DB 670 PGTKALSAVSAQAAAAQKNKLEKPEGGSFR 659

RESULT 2
US-09-724-676A-63927
; Sequence 63927, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63927
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63927

Query Match 19.5%; Score 539; DB 5; Length 709;
Best Local Similarity 30.2%; Pred. No. 3.1e-22;
Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

OY 13 PKVSSG-----PRLPAPQIVAVKAPNTTITQFPANLQLPCTGLIKNSGGLM 61
DB 286 PRATTSIGRATLTPVLAAPLPOP-----PONPTNIQ-----NQLPFGVAVLHSENGQL 336
OY 62 LVSPOQTVTR-----AETTSNITSRPVAPNPQTVKICIVPNSSSOLIKKAVATPVKKL 115
DB 337 MI-PQALAQMAQAHQAPQTTMAPRAPPTSPAPVOISVQAPGPTIAR-QVTP----- 390
OY 116 AQIGTVVTVTPKSSVQSAVAVPTSVTVTPGKPLNTVT--TLKPSLSGSS-----TPS 168
DB 391 -----TLTIKQV---SQAQITVOPSAITLQKSPGVQPOLVLGGAQOTASLGTATAVQGTGPQ 443
OY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNPLAMLIKLAGSGSQSPBEMQONVKLVEOL 227
DB 444 RYVPGATTTSSATE-----TMENVKCKNPLSTLIKLAGSGSQSTETANVKELVQNL 497
OY 228 LDKITAEETRLKLYELKSSPOHLPFLKKSVALROLPLNSOSFTQOCVOQ-----TS 283
DB 498 LDKITAEEDTSLRYELKSSPOHLPFLKKSVALROLPLNSOSFTQOCVOQ-----TS 557
OY 284 SDVVIATCTTIVTSPVTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
DB 558 Q-----ATTALTAVLVSSSVQRTAGKTAATYTSALQPVLSL----- 594
OY 344 VTLHVGPTAATGTTAGTGLQTSKPLVTSVANVTVTSLOPEKPVVSGTAVTSLPVLAV 403
DB 595 -----TQPTQVGVKQOGQPTPLVQ-----QPKP-----GALIRPQV 628
OY 404 TFGTSGAALCLPSVKPVSEFCMDHICKPVIGTPVOIKLAQPGPVLSQAPAGIPTGSSSKQ 463
DB 629 TLITQT-----PMVALRQPH-NRIMLTTPQOIQ----- 655
OY 464 LFSLFHVVOQPSGSGNEKOVYTTISHSSTLTIOKCGOKTMTPTIIPISQFPASILKOITL 523
DB 656 -----NPLQPVVVKRAVL 669
OY 524 PGKIL---SLQASPTQKNRIKENVTSCFR 550
DB 670 PGTKALSAVSAQAAAAQKNKLEKPEGGSFR 659
```

```
RESULT 3
US-09-724-676-63930
; Sequence 63930, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63930
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63930

Query Match 19.5%; Score 539; DB 5; Length 741;
Best Local Similarity 30.2%; Pred. No. 3.2e-22;
Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

OY 13 PKVSSG-----PRLPAPQIVAVKAPNTTITQFPANLQLPCTGLIKNSGGLM 61
DB 318 PRATTSIGRATLTPVLAAPLPOP-----PONPTNIQ-----NQLPFGVAVLHSENGQL 368
OY 62 LVSPOQTVTR-----AETTSNITSRPVAPNPQTVKICIVPNSSSOLIKKAVATPVKKL 115
DB 369 MI-PQALAQMAQAHQAPQTTMAPRAPPTSPAPVOISVQAPGPTIAR-QVTP----- 422
OY 116 AQIGTVVTVTPKSSVQSAVAVPTSVTVTPGKPLNTVT--TLKPSLSGSS-----TPS 168
DB 423 -----TLTIKQV---SQAQITVOPSAITLQKSPGVQPOLVLGGAQOTASLGTATAVQGTGPQ 475
OY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNPLAMLIKLAGSGSQSPBEMQONVKLVEOL 227
DB 476 RYVPGATTTSSATE-----TMENVKCKNPLSTLIKLAGSGSQSTETANVKELVQNL 529
OY 228 LDKITAEETRLKLYELKSSPOHLPFLKKSVALROLPLNSOSFTQOCVOQ-----TS 283
DB 530 LDKITAEEDTSLRYELKSSPOHLPFLKKSVALROLPLNSOSFTQOCVOQ-----TS 589
OY 284 SDVVIATCTTIVTSPVTTVSSOSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGV 343
DB 590 Q-----ATTALTAVLVSSSVQRTAGKTAATYTSALQPVLSL----- 626
OY 344 VTLHVGPTAATGTTAGTGLQTSKPLVTSVANVTVTSLOPEKPVVSGTAVTSLPVLAV 403
DB 627 -----TQPTQVGVKQOGQPTPLVQ-----QPKP-----GALIRPQV 660
OY 404 TFGTSGAALCLPSVKPVSEFCMDHICKPVIGTPVOIKLAQPGPVLSQAPAGIPTGSSSKQ 463
DB 661 TLITQT-----PMVALRQPH-NRIMLTTPQOIQ----- 667
OY 464 LFSLFHVVOQPSGSGNEKOVYTTISHSSTLTIOKCGOKTMTPTIIPISQFPASILKOITL 523
DB 688 -----NPLQPVVVKRAVL 701
OY 524 PGKIL---SLQASPTQKNRIKENVTSCFR 550
DB 702 PGTKALSAVSAQAAAAQKNKLEKPEGGSFR 731

RESULT 4
US-09-724-676A-63930
; Sequence 63930, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
```



```

Db 487 ----TTIKOV-----SOAQTTPSATLQRSBGVOPOLVLGGAQTASLGATAVQTGPQ 539
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILKACSSQSPBGMCKNKLVEOL 227
Db 540 RRVGATTTSSAATE-----TMENVKCKNFSTLIKILKACSSQSPBGMCKNKLVEOL 593
QY 228 LDKAIEAEFTKRLVYELKSSPOPFLKRSVALROLPLPNSQSFIOOCVQO-----TS 283
Db 594 LDKAIEAEFTKRLVYELKSSPOPFLKRSVALROLPLPNSQSFIOOCVQO-----TS 653
QY 284 SDMVATICTTTVTTSPTVTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
Db 654 Q-----ATTALTAVLSSSVORTAGKTAATVTSALQPPVLSL----- 690
QY 344 VTLHSGPTAAGTGTAGGILQTSKPLVTSVANTVTTVSLQPEKPVSGTAATVLSLPAV 403
Db 691 -----TQPTQVGVKQOGOPPLVLYO-----QPKP-----GALIRPPQV 724
QY 404 TFGETSGAALCLPSVVKPVVSCMDHICKPVYIGTPVOIKLAQPGPVLSQPAIGPTGSSSKQ 463
Db 725 TLTQT-----PMVALRQPH-NRIMLTTPQOIQ----- 751
QY 464 LFSLFHVQOPSGGNEKQVTTISHSSLTITQKQOKTNPVNTIIPISQFPASILKQITL 523
Db 752 -----NPLQPPVVPVAVL 765
QY 524 PGNKIL-----SLQASPTQKNRIKENVTSQCFR 550
Db 766 PGTKALSAVSAQAAMAKNKLKEPBGGSFR 795

RESULT 7
US-09-724-676-63924
; Sequence 63924, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63924
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63924

```

Query Match 19.5%; Score 539; DB 5; Length 837;
 Best Local Similarity 30.2%; Pred. No. 3,7e-22;
 Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

```

QY 13 PRVSSG-----PRLPAQIYAVAKAPMTTITQFPANLQLPFGVILKINSQPLM 61
Db 414 PRATTSIGIRATLPTVLAIRLPQ-----PQNPNTIQ--NFQLPBGVILVSENGQL 464
QY 62 LVSPOQTVR-----AETTSNITSRPVAPNQTIVKICIVPNSSQOLIKKAVTPVKKL 115
Db 465 MI-PQALAQMOQAHAQOTTMARPRATPTSAAPVOISTVQAPGPIIAR-QVTP----- 518
QY 116 AQIGTIVTTVPKPSVQSAVPTSVYTVTPGKPLNTVT--TLKPSISLASS-----TPS 168
Db 519 -----TTIIQV--SOAQTTPSATLQRSBGVOPOLVLGGAQTASLGATAVQTGPQ 571
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILKACSSQSPBGMCKNKLVEOL 227
Db 572 RRVGATTTSSAATE-----TMENVKCKNFSTLIKILKACSSQSPBGMCKNKLVEOL 593
QY 228 LDKAIEAEFTKRLVYELKSSPOPFLKRSVALROLPLPNSQSFIOOCVQO-----TS 283
Db 626 LDKAIEAEFTKRLVYELKSSPOPFLKRSVALROLPLPNSQSFIOOCVQO-----TS 653
QY 284 SDMVATICTTTVTTSPTVTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343

```

```

Db 686 Q-----ATTALTAVLSSSVORTAGKTAATVTSALQPPVLSL----- 722
QY 344 VTLHSGPTAAGTGTAGGILQTSKPLVTSVANTVTTVSLQPEKPVSGTAATVLSLPAV 403
Db 723 -----TQPTQVGVKQOGOPPLVLYO-----QPKP-----GALIRPPQV 756
QY 404 TFGETSGAALCLPSVVKPVVSCMDHICKPVYIGTPVOIKLAQPGPVLSQPAIGPTGSSSKQ 463
Db 757 TLTQT-----PMVALRQPH-NRIMLTTPQOIQ----- 783
QY 464 LFSLFHVQOPSGGNEKQVTTISHSSLTITQKQOKTNPVNTIIPISQFPASILKQITL 523
Db 784 -----NPLQPPVVPVAVL 797
QY 524 PGNKIL-----SLQASPTQKNRIKENVTSQCFR 550
Db 798 PGTKALSAVSAQAAMAKNKLKEPBGGSFR 827

RESULT 8
US-09-724-676A-63924
; Sequence 63924, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63924
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63924

```

Query Match 19.5%; Score 539; DB 5; Length 837;
 Best Local Similarity 30.2%; Pred. No. 3,7e-22;
 Matches 172; Conservative 66; Mismatches 144; Indels 188; Gaps 21;

```

QY 13 PRVSSG-----PRLPAQIYAVAKAPMTTITQFPANLQLPFGVILKINSQPLM 61
Db 414 PRATTSIGIRATLPTVLAIRLPQ-----PQNPNTIQ--NFQLPBGVILVSENGQL 464
QY 62 LVSPOQTVR-----AETTSNITSRPVAPNQTIVKICIVPNSSQOLIKKAVTPVKKL 115
Db 465 MI-PQALAQMOQAHAQOTTMARPRATPTSAAPVOISTVQAPGPIIAR-QVTP----- 518
QY 116 AQIGTIVTTVPKPSVQSAVPTSVYTVTPGKPLNTVT--TLKPSISLASS-----TPS 168
Db 519 -----TTIKOV--SOAQTTPSATLQRSBGVOPOLVLGGAQTASLGATAVQTGPQ 571
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILKACSSQSPBGMCKNKLVEOL 227
Db 572 RRVGATTTSSAATE-----TMENVKCKNFSTLIKILKACSSQSPBGMCKNKLVEOL 593
QY 228 LDKAIEAEFTKRLVYELKSSPOPFLKRSVALROLPLPNSQSFIOOCVQO-----TS 283
Db 626 LDKAIEAEFTKRLVYELKSSPOPFLKRSVALROLPLPNSQSFIOOCVQO-----TS 653
QY 284 SDMVATICTTTVTTSPTVTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGVGAAGV 343
Db 686 Q-----ATTALTAVLSSSVORTAGKTAATVTSALQPPVLSL----- 722
QY 344 VTLHSGPTAAGTGTAGGILQTSKPLVTSVANTVTTVSLQPEKPVSGTAATVLSLPAV 403
Db 723 -----TQPTQVGVKQOGOPPLVLYO-----QPKP-----GALIRPPQV 756
QY 404 TFGETSGAALCLPSVVKPVVSCMDHICKPVYIGTPVOIKLAQPGPVLSQPAIGPTGSSSKQ 463
Db 757 TLTQT-----PMVALRQPH-NRIMLTTPQOIQ----- 783

```


QY 464 LFSLEHVVOQPSGNGENKQVTTISHSSTLTIQCGQKTMPTVTTIPTSOPPASILKQITL 523
Db 784 -----NPLQPVVPPVKKPAVL 797
QY 524 GKNKIL---SLQASPTOKNRKIKENTSCFR 550
Db 798 GTRKLSAVSAQAQAAAKNKLKEPGGSFR 827

RESULT 9

US-09-724-676-63928
; Sequence 63928, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63928
; LENGTH: 925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63928

Query Match 19.1%; Score 527; DB 5; Length 925;
Best Local Similarity 30.6%; Pred. No. 1,9e-21;

Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPCTVLIKNSGPI 61
Db 286 PRTATSGIRATLPTVLAFLRPOP-----PQNTNIO---NFQLPFGMLVRSNGQL 336
QY 62 LVSPQQTVTR-----AETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATPVK 115
Db 337 MI-POQALQMOQAOHAOPQTTMAPRPTSPAPVOISTVQAGCTPIAR-QVTP----- 390
QY 116 AQTGTTVTTVPKSSVQSAVAVPTSVTVTPGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 391 -----TTTIKQV---SQAOITVQPSATLQSRSPGVPOLVLSGAQTAISLGTATVOTGPQ 443
QY 169 NE-PNLKAENSAVAQINLSPTMLENVKCKNFMLIKLACSGSOSPDMGONVKKLEVOL 227
Db 444 RTVGATTTSSAATE-----TMENVKCKNFSTLIKILASSGOSTETANVKELYONL 497
QY 228 LDKAIEEFTRLKLYELKSSPOPHLVPLKKSVAVALRQLLPNSQSTIQOCVQOTSSDMV 287
Db 498 LDKRTEADEFTSRLYRELNSSPOPHLVPLKRSPLALROLPPDSAAFIOQSOQOPPP-- 555
QY 288 IATCTTIVTSPVTTVSSSSQSEKSIIVSGATPRTVSVOGTLNPLAGPVGAKAGVYTLH 347
Db 556 -----TSQATTA--LTAVVLS-----SVQ-----R 574
QY 348 SVGPPTAATGCTAGTGLQTSKPLTVSVANTVTTVSLQPEKPVVSGTAVTLSPAVTGE 407
Db 575 TAGKTAAT-----VTSALQP--PVSLT----- 595
QY 408 TSGAICLPVSKPVVSCMDHICKRPVIGTPVQI--KLAOPGVLSQPAIGPTSSSKOL 464
Db 596 -----OPTYGVGKOGQPTPL----- 611
QY 465 FSLFHVVOQPSGNGENKQVTTISHSSTLTIQCGQKTMPTVTTIPTSOPPASILKQITL 524
Db 612 -----VIOQP-----PKPG-----ALIDLNLPLOPVVKKPAVL 640
QY 525 GKNKIL---SLQASPTOKNRKIKENTSCFRDE 552
Db 641 GTRKLSAVSAQAQAAAKNKLKEPGGSFRD 671

RESULT 10
US-09-724-676A-63928

; Sequence 63928, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63928
; LENGTH: 925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63928

Query Match 19.1%; Score 527; DB 5; Length 925;
Best Local Similarity 30.6%; Pred. No. 1,9e-21;

Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;

QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPCTVLIKNSGPI 61
Db 286 PRTATSGIRATLPTVLAFLRPOP-----PQNTNIO---NFQLPFGMLVRSNGQL 336
QY 62 LVSPQQTVTR-----AETTSNITSRAVPANPQTVKICVTPNSSSOLIKKVAATPVK 115
Db 337 MI-POQALQMOQAOHAOPQTTMAPRPTSPAPVOISTVQAGCTPIAR-QVTP----- 390
QY 116 AQTGTTVTTVPKSSVQSAVAVPTSVTVTPGKPLNTVT--TLKPSLSGASS-----TPS 168
Db 391 -----TTTIKQV---SQAOITVQPSATLQSRSPGVPOLVLSGAQTAISLGTATVOTGPQ 443
QY 169 NE-PNLKAENSAVAQINLSPTMLENVKCKNFMLIKLACSGSOSPDMGONVKKLEVOL 227
Db 444 RTVGATTTSSAATE-----TMENVKCKNFSTLIKILASSGOSTETANVKELYONL 497
QY 228 LDKAIEEFTRLKLYELKSSPOPHLVPLKKSVAVALRQLLPNSQSTIQOCVQOTSSDMV 287
Db 498 LDKRTEADEFTSRLYRELNSSPOPHLVPLKRSPLALROLPPDSAAFIOQSOQOPPP-- 555
QY 288 IATCTTIVTSPVTTVSSSSQSEKSIIVSGATPRTVSVOGTLNPLAGPVGAKAGVYTLH 347
Db 556 -----TSQATTA--LTAVVLS-----SVQ-----R 574
QY 348 SVGPPTAATGCTAGTGLQTSKPLTVSVANTVTTVSLQPEKPVVSGTAVTLSPAVTGE 407
Db 575 TAGKTAAT-----VTSALQP--PVSLT----- 595
QY 408 TSGAICLPVSKPVVSCMDHICKRPVIGTPVQI--KLAOPGVLSQPAIGPTSSSKOL 464
Db 596 -----OPTYGVGKOGQPTPL----- 611
QY 465 FSLFHVVOQPSGNGENKQVTTISHSSTLTIQCGQKTMPTVTTIPTSOPPASILKQITL 524
Db 612 -----VIOQP-----PKPG-----ALIDLNLPLOPVVKKPAVL 640
QY 525 GKNKIL---SLQASPTOKNRKIKENTSCFRDE 552
Db 641 GTRKLSAVSAQAQAAAKNKLKEPGGSFRD 671

RESULT 11
US-09-724-676-63931
; Sequence 63931, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63931
; LENGTH: 957

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63931

Query Match
Best Local Similarity 30.6%; Pred. No. 26-21;
Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;

QY 13 PKRVSSG-----PRAPQIVAAKAPRTTITQPPANIQLPQGVYLIKNSGPI 61
DB 318 PRATTSIGIRATLPTVLARLP-----PONPTNIQ--NFQLPGWLVASENGOLL 368
QY 62 LVSPQOQTVTR-----AETTSNITSRPAPVAPNQTIVKICTPVNSSQLIKKVAVTPVK 115
DB 369 MI-POQALQMOQAHAQOQTMAPRPTSPAPVOISTVQAPGPPIIAR--QVTP----- 422
QY 116 AOGITVVTTPKRVSSVQSAVPTSVVTPGKPLNTVT--TLKPSLIGASS-----TPS 168
DB 423 ----TIIKOV---SOAQTTVQPSATLQSPGVQPOLVIGGAQOTASLGTATAVQGTQ 475
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILACSGSGSPENGQNVKLI 227
DB 476 RTVPQATTTSSATE-----TMENVKCKNFSTLIKILASSGOSTETANVKEIVONL 529
QY 228 LDKITAEFTTRKLYELKSSPOPHLVPLKKSVALRQLPNSQSFIOQCVOQTSSDW 287
DB 530 LDKITAEFTSRKLYELKSSPOPHLVPLKKSVALRQLPNSQSFIOQCVOQTSSDW 587
QY 288 IATCTTTVTPSVVTTTVSSQSEKSIIVGATAPRTVSQTLNPLAGPVGAKAGVYTLH 347
DB 588 ----TSQATTA--LTAVALSS-----SVQ-----R 606
QY 348 SVGPTATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPVTGE 407
DB 607 TAGKTAAT-----VTSALQ--PVLSTL----- 627
QY 408 TSGAATCLPSVKRVSVFCMDHICKPVIGTPVQI---KLAQPGVLSQAPACIPGSSSKOL 464
DB 628 ----VIQDP-----PKPG-----ALIQLPNLPQVPVAKVAVLP 672
QY 465 FSLFHVQOPSGNGEKQVTTISHSSTLTIOCKGQKTMPTVTIIPTSQFPASILKQITLP 524
DB 644 ----VIQDP-----PKPG-----ALIQLPNLPQVPVAKVAVLP 672
QY 525 GKKIL---SLOASPTQKNRIKENVTSCFRDE 552
DB 673 GTKALSAVSAQAAQAQKNKIKKEPGGGSFRD 703

RESULT 12
US-09-724-676A-63931
; Sequence 63931, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63931
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63931

Query Match
Best Local Similarity 19.1%; Score 527; DB 5; Length 957;
Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;

QY 13 PKRVSSG-----PRAPQIVAAKAPRTTITQPPANIQLPQGVYLIKNSGPI 61
DB 318 PRATTSIGIRATLPTVLARLP-----PONPTNIQ--NFQLPGWLVASENGOLL 368
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILACSGSGSPENGQNVKLI 227
DB 476 RTVPQATTTSSATE-----TMENVKCKNFSTLIKILASSGOSTETANVKEIVONL 529
QY 228 LDKITAEFTTRKLYELKSSPOPHLVPLKKSVALRQLPNSQSFIOQCVOQTSSDW 287
DB 530 LDKITAEFTSRKLYELKSSPOPHLVPLKKSVALRQLPNSQSFIOQCVOQTSSDW 587
QY 288 IATCTTTVTPSVVTTTVSSQSEKSIIVGATAPRTVSQTLNPLAGPVGAKAGVYTLH 347
DB 588 ----TSQATTA--LTAVALSS-----SVQ-----R 606
QY 348 SVGPTATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPVTGE 407
DB 607 TAGKTAAT-----VTSALQ--PVLSTL----- 627
QY 408 TSGAATCLPSVKRVSVFCMDHICKPVIGTPVQI---KLAQPGVLSQAPACIPGSSSKOL 464
DB 628 ----VIQDP-----PKPG-----ALIQLPNLPQVPVAKVAVLP 672
QY 465 FSLFHVQOPSGNGEKQVTTISHSSTLTIOCKGQKTMPTVTIIPTSQFPASILKQITLP 524
DB 644 ----VIQDP-----PKPG-----ALIQLPNLPQVPVAKVAVLP 672
QY 525 GKKIL---SLOASPTQKNRIKENVTSCFRDE 552
DB 673 GTKALSAVSAQAAQAQKNKIKKEPGGGSFRD 703

```

```

QY 62 LVSPQOQTVTR-----AETTSNITSRPAPVAPNQTIVKICTPVNSSQLIKKVAVTPVK 115
DB 369 MI-POQALQMOQAHAQOQTMAPRPTSPAPVOISTVQAPGPPIIAR--QVTP----- 422
QY 116 AOGITVVTTPKRVSSVQSAVPTSVVTPGKPLNTVT--TLKPSLIGASS-----TPS 168
DB 423 ----TIIKOV---SOAQTTVQPSATLQSPGVQPOLVIGGAQOTASLGTATAVQGTQ 475
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILACSGSGSPENGQNVKLI 227
DB 476 RTVPQATTTSSATE-----TMENVKCKNFSTLIKILASSGOSTETANVKEIVONL 529
QY 228 LDKITAEFTTRKLYELKSSPOPHLVPLKKSVALRQLPNSQSFIOQCVOQTSSDW 287
DB 530 LDKITAEFTSRKLYELKSSPOPHLVPLKKSVALRQLPNSQSFIOQCVOQTSSDW 587
QY 288 IATCTTTVTPSVVTTTVSSQSEKSIIVGATAPRTVSQTLNPLAGPVGAKAGVYTLH 347
DB 588 ----TSQATTA--LTAVALSS-----SVQ-----R 606
QY 348 SVGPTATGTTAGTGLQTSKPLVTSVANTVTVSLOPEKPVSGTAVTSLPVTGE 407
DB 607 TAGKTAAT-----VTSALQ--PVLSTL----- 627
QY 408 TSGAATCLPSVKRVSVFCMDHICKPVIGTPVQI---KLAQPGVLSQAPACIPGSSSKOL 464
DB 628 ----VIQDP-----PKPG-----ALIQLPNLPQVPVAKVAVLP 672
QY 465 FSLFHVQOPSGNGEKQVTTISHSSTLTIOCKGQKTMPTVTIIPTSQFPASILKQITLP 524
DB 644 ----VIQDP-----PKPG-----ALIQLPNLPQVPVAKVAVLP 672
QY 525 GKKIL---SLOASPTQKNRIKENVTSCFRDE 552
DB 673 GTKALSAVSAQAAQAQKNKIKKEPGGGSFRD 703

RESULT 13
US-09-724-676-63922
; Sequence 63922, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63922
; LENGTH: 1021
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63922

Query Match
Best Local Similarity 19.1%; Score 527; DB 5; Length 1021;
Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;

QY 13 PKRVSSG-----PRAPQIVAAKAPRTTITQPPANIQLPQGVYLIKNSGPI 61
DB 382 PRATTSIGIRATLPTVLARLP-----PONPTNIQ--NFQLPGWLVASENGOLL 432
QY 62 LVSPQOQTVTR-----AETTSNITSRPAPVAPNQTIVKICTPVNSSQLIKKVAVTPVK 115
DB 433 MI-POQALQMOQAHAQOQTMAPRPTSPAPVOISTVQAPGPPIIAR--QVTP----- 486
QY 116 AOGITVVTTPKRVSSVQSAVPTSVVTPGKPLNTVT--TLKPSLIGASS-----TPS 168
DB 487 ----TIIKOV---SOAQTTVQPSATLQSPGVQPOLVIGGAQOTASLGTATAVQGTQ 539
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKILACSGSGSPENGQNVKLI 227

```

Db 540 RTVPGATTTSSAATE-----TMENVKCKNFSLTLIKLASSGKOSTETANVKELYONL 593
Qy 228 LDKIAEEFTFKLYVELKSSPOPLVLPFLKKSVALRQLLPNSQSFIOQCVOQTSSIMV 287
Db 594 LDGKIEADEFTSRILRELNSSPOPLVLPFLKRSPLALRQLLPDSAAFIOQSOQOQPPPP-- 651
Qy 288 IATCTTATTSPPVTTTSSQSEKSIIVSGATAPRTVSVOTLNLAPGVAKAGAVTLH 347
Db 652 ----TSQATTA--LTAVALSS-----SVQ-----R 670
Qy 348 SVGPPTAAGCTTAGGLQTSKPLVTSVANTVTYVLSLOPEKPVVSGTAVTSLPAVTEGE 407
Db 671 TAGKTAAT-----VTSALDP--PVLSLT----- 691
Qy 408 TSGAAILCPVSKPVVSPFCMDHICKPVIGTVPVQI---KLAQPGPVLSQAPAGIPTGSSSKQL 464
Db 692 ----TSQATTA--LTAVALSS-----SVQ-----R 707
Qy 465 FSLFHVVOQPSGNEKQVTTTSHSSTLTIOKCGQKTMVNTIIPTSQPPASILKQITLP 524
Db 708 ----VIOQP-----PKPG-----ALIQNLPLQPPVVPVAVPLP 736
Qy 525 GNKIL---SLOASPTOKRIKENVTSCFRDE 552
Db 737 GTKALSAVSAQAAAAQKNKKEPGGGSFRDD 767
RESULT 14
US-09-724-676A-63922
Sequence 63922, Application US/09724676A
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724, 676A
NUMBER OF FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63922
LENGTH: 1021
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676A-63922
Query Match 19.1%; Score 527; DB 5; Length 1021;
Best Local Similarity 30.6%; Pred. No. 2.1e-21;
Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;
Qy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPFGTVLKSNSGPLM 61
Db 382 PTATTSIGRATLTPTVLAPRLPQ-----PQNPTNIO--NFQLPQGMVLVRSNGQLL 432
Qy 62 LVSPQOQVTR-----AETTSNITSRPVAVPANPQTVKICTVPPNSSQLIKKVAVTPVKKL 115
Db 433 MI-POQALAOQAOAHAPQOTTMARPRPTPSAPPVQISTVOAGETPIIAR-QVTP----- 486
Qy 116 AIGTIVTTVPKPSVSVAVPVSIVVTGPKPLNTVT--TLKPSISGASS-----TPS 168
Db 487 ----TTIKQV---SQAOITVQPSATLQSPGVQPOLVLGAAQASLGTATNAVOTGTPQ 539
Qy 169 NE-PRUKAENSAAVOINISPTMLENVKCKNFAMLIKLAGSGSPPEMGONVKVLQOL 227
Db 540 RTVPGATTTSSAATE-----TMENVKCKNFSLTLIKLASSGKOSTETANVKELYONL 593
Qy 228 LDKIAEEFTFKLYVELKSSPOPLVLPFLKKSVALRQLLPNSQSFIOQCVOQTSSIMV 287
Db 594 LDGKIEADEFTSRILRELNSSPOPLVLPFLKRSPLALRQLLPDSAAFIOQSOQOQPPPP-- 651
Qy 288 IATCTTATTSPPVTTTSSQSEKSIIVSGATAPRTVSVOTLNLAPGVAKAGAVTLH 347
Db 652 ----TSQATTA--LTAVALSS-----SVQ-----R 670
Qy 348 SVGPPTAAGCTTAGGLQTSKPLVTSVANTVTYVLSLOPEKPVVSGTAVTSLPAVTEGE 407

Db 671 TAGKTAAT-----VTSALDP--PVLSLT----- 691
Qy 408 TSGAAILCPVSKPVVSPFCMDHICKPVIGTVPVQI---KLAQPGPVLSQAPAGIPTGSSSKQL 464
Db 692 ----TSQATTA--LTAVALSS-----SVQ-----R 707
Qy 465 FSLFHVVOQPSGNEKQVTTTSHSSTLTIOKCGQKTMVNTIIPTSQPPASILKQITLP 524
Db 708 ----VIOQP-----PKPG-----ALIQNLPLQPPVVPVAVPLP 736
Qy 525 GNKIL---SLOASPTOKRIKENVTSCFRDE 552
Db 737 GTKALSAVSAQAAAAQKNKKEPGGGSFRDD 767
RESULT 15
US-09-724-676-63925
Sequence 63925, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724, 676
NUMBER OF FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63925
LENGTH: 1053
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-63925
Query Match 19.1%; Score 527; DB 5; Length 1053;
Best Local Similarity 30.6%; Pred. No. 2.2e-21;
Matches 175; Conservative 58; Mismatches 122; Indels 216; Gaps 24;
Qy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPFGTVLKSNSGPLM 61
Db 414 PTATTSIGRATLTPTVLAPRLPQ-----PQNPTNIO--NFQLPQGMVLVRSNGQLL 464
Qy 62 LVSPQOQVTR-----AETTSNITSRPVAVPANPQTVKICTVPPNSSQLIKKVAVTPVKKL 115
Db 465 MI-POQALAOQAOAHAPQOTTMARPRPTPSAPPVQISTVOAGETPIIAR-QVTP----- 518
Qy 116 AIGTIVTTVPKPSVSVAVPVSIVVTGPKPLNTVT--TLKPSISGASS-----TPS 168
Db 519 ----TTIKQV---SQAOITVQPSATLQSPGVQPOLVLGAAQASLGTATNAVOTGTPQ 571
Qy 169 NE-PRUKAENSAAVOINISPTMLENVKCKNFAMLIKLAGSGSPPEMGONVKVLQOL 227
Db 572 RTVPGATTTSSAATE-----TMENVKCKNFSLTLIKLASSGKOSTETANVKELYONL 625
Qy 228 LDKIAEEFTFKLYVELKSSPOPLVLPFLKKSVALRQLLPNSQSFIOQCVOQTSSIMV 287
Db 626 LDGKIEADEFTSRILRELNSSPOPLVLPFLKRSPLALRQLLPDSAAFIOQSOQOQPPPP-- 683
Qy 288 IATCTTATTSPPVTTTSSQSEKSIIVSGATAPRTVSVOTLNLAPGVAKAGAVTLH 347
Db 684 ----TSQATTA--LTAVALSS-----SVQ-----R 702
Qy 348 SVGPPTAAGCTTAGGLQTSKPLVTSVANTVTYVLSLOPEKPVVSGTAVTSLPAVTEGE 407
Db 703 TAGKTAAT-----VTSALDP--PVLSLT----- 723
Qy 408 TSGAAILCPVSKPVVSPFCMDHICKPVIGTVPVQI---KLAQPGPVLSQAPAGIPTGSSSKQL 464
Db 724 ----TSQATTA--LTAVALSS-----SVQ-----R 739
Qy 465 FSLFHVVOQPSGNEKQVTTTSHSSTLTIOKCGQKTMVNTIIPTSQPPASILKQITLP 524
Db 740 ----VIOQP-----PKPG-----ALIQNLPLQPPVVPVAVPLP 768

Thu Feb 20 16:36:34 2003

us-09-763-909-2_copy_1_552.rapn

Page 8

```

Oy      525 GNKIL--SLQASPTQKNRIKENTVSCFRDE 552
          | | | | | | | | | | | | | |
Db      769 GTTALSAVSAQAAAAQKNKLTKEPGGSFRDD 799

```

Search completed: February 16, 2003, 22:04:20
Job time : 30.4582 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 18.6758 Seconds
(Without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GLLVTKVAPVSAAPPKVSQSP.....ASPTQKNRIKENVTSCFRDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	256	9.3	921	2 A48184	transcription init
2	256	9.3	921	2 A45183	TBP-associated fac
3	234.5	8.5	3020	2 A43932	mucin 2 precursor,
4	233.5	8.5	528	2 I47141	gastric mucin (clo
5	226	8.2	1275	2 T33369	hypothetical prote
6	221	8.0	2035	2 A40718	host cell factor C
7	220	8.0	1367	1 S48478	glucan 1,4-alpha-g
8	212.5	7.7	2232	2 T34434	hypothetical prote
9	210	7.6	873	2 A47283	calphostin - fruit
10	208	7.5	1777	2 T34369	hypothetical prote
11	206.5	7.5	865	2 A47282	calcium-binding pr
12	205	7.4	1032	2 T34433	hypothetical prote
13	204.5	7.4	1324	2 S52863	DNA-binding protei
14	204	7.4	2187	2 T30826	nascant polyepitid
15	201	7.3	873	2 P96615	probable Myb-famil
16	197.5	7.2	5376	2 T42215	zonahesin - mouse
17	197	7.1	648	2 PC4395	mucin 3 - human (f
18	196.5	7.1	3507	2 T45463	membrane glycoprot
19	196	7.1	3507	2 T34513	hypothetical prote
20	195	7.1	1161	2 S57180	probable membrane
21	193.5	7.0	670	2 P36791	hypothetical prote
22	193	7.0	825	2 T29634	hypothetical prote
23	192	6.9	1367	2 S51959	hypothetical prote
24	189.5	6.9	797	1 VGBEX1	glycoprotein x pre
25	188	6.8	866	2 T45462	membrane glycoprot
26	188	6.8	3570	2 T45025	mucin MUC5B, trach
27	186	6.7	1151	2 T18535	high molecular mas
28	185.5	6.7	1630	2 A53577	ascites stialoglyco
29	184.5	6.7	881	2 S56032	probable membrane

30	181.5	6.6	662	2 A45155	mucin FTM-C.1 - Af
31	181.5	6.6	1199	2 A40670	nuclear envelope p
32	181	6.6	1609	2 S25345	probable membrane
33	180.5	6.5	725	2 A41258	a-aggutinin core
34	179.5	6.5	1537	2 S53465	fiocculatation prote
35	176.5	6.4	1260	2 S60896	agglutinin-like pr
36	176.5	6.4	1851	2 T119964	hypothetical prote
37	175.5	6.4	796	2 T21460	hypothetical prote
38	175.5	6.4	2090	2 S26058	probable transform
39	174	6.3	549	2 C87719	protein R119.6 (im
40	174	6.3	1075	2 S48992	fiocculatation prote
41	173.5	6.3	583	2 S67571	hypothetical prote
42	172.5	6.3	851	2 T22696	hypothetical prote
43	172.5	6.3	1858	2 T18273	1-phosphatidylinos
44	172	6.2	886	2 S29605	glycoprotein 350/2
45	172	6.2	1169	2 S38181	fiocculatation prote

ALIGNMENTS

RESULT 1

A48184
transcription initiation factor IID 110k chain - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A48184
R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 5896-5900, 1993
A:Title: The Drosophila 110-kDa transcription factor TFIID subunit directly interacts
A:Reference number: A48184; MUID:93171591; PMID:8327460
A:Accession: A48184
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-921 <KOK>
A:Cross-references: GB:563550; NID:9398432; PID:9398433
A:Experimental source: embryo nuclear extract
A>Note: sequence extracted from NCHI backbone (NCBIN:134863, NCBI:134864)
C:Genetics:
A:Gene: FlyBase:Trf110
A:Cross-references: FlyBase:FBgn0010280
C:Keywords: transcription Initiation

Query Match	9.3%;	Score 256;	DB 2;	Length 921;
Best Local Similarity	22.9%;	Pred. No. 3.8e-07;		
Matches 135;	Conservative 86;	Mismatches 193;	Indels 176;	Gaps 24;
QY	25	POIYAVKAPNTTQIOPANQLPQTVLIRKNSGFLML-----VSPQQTVAETTS	76	
DB	133	POSPSITLSTLTNGOTPA-----LLVKTNDNGFQLLRVGTGPPVLTQITNTSNS	184	
QY	77	NITSPPANPAPQVYKICTVNSSQ-----LIKVAVTPPKKLAQIGTVYTVIP	127	
DB	185	NTTSSTNNHTTQO-IRLOTVPAASMTNTTATSMIIVASVSGVANSQPHPLTOLNAQ	243	
QY	128	KPSSQSAVAPVSVTVTPGKPLNTVTTLKPSISGASSTPNEPLKKAENSAVQINLSP	187	
DB	244	APQLPQIOTITPAQSQQQQVNNVSSAGTATRAVSSTTA-----ATT	287	
QY	188	TMLENVR-KCKNFLMLIKLACSGSQSPKMGONKVLQDLDAKIEAEFTKRLYEELK	246	
DB	288	TOOGTKKCKRFLANLIEL--SPREKPVKKNVNTLQELVANNVDEEFCDRLERLN	345	
QY	247	SSPQHLVPLFKKSVALROL-----LPSQSTQ---	276	
DB	346	ASPPCLIGFLFKSLPLRLQALYKELVIEGIKPPQHVGLAGLSQQLPIQAOIRDIG	405	
QY	277	---QCVQOTSDDVIATCTTVTTSPPVTTVSSQSEKSLIISGATAPRTVS---VQT	329	
DB	406	PSQTTTIGQTVYRMI--TPNALGTPRPRTIGTTTSKQPN---IRLPAPRLVNTGGIRT	460	
QY	330	LNPLAGPVGARAVVTLHSVGPRTAATGTTAGTGLQTSKPLVTSVANTVTVTSIQPEPK	389	

```

Db 461 QIP-SLOVPGQANIVQIR--GPOHOLQRTGSVQIRATRRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPRAVTEGERSGAALICLPKPVVSVFCMDHCKRPVITGPVQIKLAQEPVL 449
Db 507 -----KLTAVKVGQFOIKAI--TPSLHP-----PSL 530
QY 450 SOPAGIPTGSSSKQLSLFHHVVOQPSGNEKQVTTISHSSTLTIOKCGQKTMPTVNTIIP 509
Db 531 AATSGPPTPTLSVLS-----TINSAST-----TTLEIPPS-LPT 564
QY 510 SQPPASIT---LKQITLPGN-----KILSLQA---SPTQKNRIKENVT 546
Db 565 VHLPPALRAREQMONSLNHNHFDKALVEIKAPSLHPHMERINASTLT 614

RESULT 2
A:Associated factor TFIIID - fruit fly (Drosophila sp.)
C:Species: Drosophila sp.
C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997
C:Accession: A45183
R:Hoey, T.; Weinzierl, R.O.; Gill, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R.
Cell 72, 247-260, 1993
A:Title: Molecular cloning and functional analysis of Drosophila TAF110 reveal properties
A:Reference number: A45183; MUID:93145326; PMID:7678780
A:Accession: A45183
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-921 <HOE>
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIP:123832)
C:Genetics:
A:Gene: FlyBase:Tafl10
A:Cross-References: FlyBase:FBgn0010280

Query Match 9.3%; Score 256; DB 2; Length 921;
Best Local Similarity 22.9%; Pred. No. 3.8e-07;
Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

QY 25 POLVAVKANTTTTIOFPANLDPGTVLTKNSGML-----VSPQQTVAETTS 76
Db 133 POSPSTLSTLNTGTPA-----LAVKTDGFPOLLKRGTTTGPVYQTITNTSNN 184
QY 77 NTSRAVPANPQTVKICVNPSSQ-----LIKKAVVAPVKKLAQIGTTVTTPV 127
Db 185 NNTSTTNTHTTQ-IRLQTVPAASMTNTTANSITIVNSVASSGVANSSOPHLLTOLMNO 243
QY 128 KSSVQSVAVPTSVYVTPGKPLNTVTLKPSLSGASSTPSNEPNLKAENSAVAQINLSP 187
Db 244 AQLPQITQITPAQOQSQOQVNNVSSAGTATAVSTTA-----ATT 287
QY 188 TMLNENK-KCKFLMLILACSGSPKMGQVKKVLQQLDAKLEAEFTKKLYEIK 246
Db 288 TQGGNTKCKRCFLANLIEL--STREPKVEKNVRLIOLVNAVNEPEFCRLRLRLN 345
QY 247 SSPQHLVFLKKSVALROL-----LPNSQSFIO----- 276
Db 346 ASPQCLIFLKKSLPLRLQALYTKELVTEGKPPQAHVLAGLSQQLKTOAQIRPIG 405
QY 277 ---OCVOQTSDDVYATCTTNTTSPVYTTVSSQSEKSIIVSGATAPRVSV---VOT 329
Db 406 PSQTTTIGTQYRMT--TPNALGTPTTGHITTSKQPN---IRLPTAPRLVNTGINT 460
QY 330 LNPPLAGPVAKAGVYTLHSVGPATAGTTAGTGLQTSKPLVTSVANTVTVVSLQPEK 389
Db 461 QIP-SLOVPGQANIVQIR--GPOHOLQRTGSVQIRATRRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPRAVTEGERSGAALICLPKPVVSVFCMDHCKRPVITGPVQIKLAQEPVL 449
Db 507 -----KLTAVKVGQFOIKAI--TPSLHP-----PSL 530
QY 450 SOPAGIPTGSSSKQLSLFHHVVOQPSGNEKQVTTISHSSTLTIOKCGQKTMPTVNTIIP 509

```

```

Db 531 AATSGPPTPTLSVLS-----TINSAST-----TTLEIPPS-LPT 564
QY 510 SQPPASIT---LKQITLPGN-----KILSLQA---SPTQKNRIKENVT 546
Db 565 VHLPPALRAREQMONSLNHNHFDKALVEIKAPSLHPHMERINASTLT 614

RESULT 3
A43932
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C>Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C:Accession: A49963; A45106; B45106; B45106; B3532; A61257; PQ0328; PQ0329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MUID:94132002; PMID:8300571
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-References: GB:M94131; NID:9186395; PIDN:AAA59103.1; PID:9186396
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A:Reference number: A45106; MUID:93016075; PMID:1400449
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-References: GB:M94131; NID:9186395; PIDN:AAA59103.1; PID:9186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-References: GB:M94132; NID:9186397; PIDN:AAA59104.1; PID:9186398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:91558717; PMID:1885763
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-References: GB:M74027; NID:9188863; PIDN:AAA59875.1; PID:9188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:55749; NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl
A:Reference number: A43532; MUID:89197956; PMID:2703501
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-References: GB:M22405; NID:9188873; PIDN:AAA6334.1; PID:9188874
A:Experimental source: Intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481; PMID:1985113
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A:Experimental source: bronchus
R:Xu, G.; Huan, L.; Khatri, T.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstn
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the
A:Reference number: PQ0328; MUID:92198477; PMID:1550588
A:Accession: PQ0328
A:Molecule type: mRNA

```


glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)
N/Alternate names: extracellular glucanase; mucin-like protein MUC1; protein YIR019c
C/Species: *Saccharomyces cerevisiae*
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 12-Nov-1999
C/Accession: S48478; A26877; B26877; S27281; J06123
R/Rowley, K.
submitted to the EMBL Data Library, October 1994
A/Reference number: S48478
A/Accession: S48478
A/Molecule type: DNA
A/Residues: 1-1367 <ROW>
A/Cross-references: GB:247047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009; MIM
J.Ryamashta, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A/Title: Gene fusion is a possible mechanism underlying the evolution of SPAL.
A/Reference number: A91831; MUID:87194600; PMID:3106330
A/Accession: A26877
A/Molecule type: DNA
A/Residues: 1-242 <YAN>
A/Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA5014.1; PID:g172525
A/Accession: B26877
A/Molecule type: DNA
A/Residues: 762-1331 <YA2>
A/Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA5015.1; PID:g172526
R.Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A/Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Sacchar*
A/Reference number: S27281; MUID:89031230; PMID:3141213
A/Accession: S27281
A/Molecule type: DNA
A/Residues: 1-31 <PAR>
A/Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R.Liambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A/Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A/Reference number: J06123; MUID:96323237; PMID:8710886
A/Accession: J06123
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-1367 <LAM>
A/Cross-references: GB:U00626; NID:g1304386; PIDN:AA049609.1; PID:g1304387
C/Genetics:
A/Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A/Cross-references: MIPS:YIR019c; SGD:S0001458
A/Map position: 9R
C/Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C/Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F/5-1/Domain: transmembrane #status predicted <TM>
F/1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 8.0% Score 220; DR 1; Length 1367;
Best Local Similarity 22.5% Pred. No. 7; Mismatches 240; Indels 134; Gaps 22;
Matches 131; Conservative 77;

2 TLTVKVAEVSAPPKVSGRLPAP-----QIVAKPANTTTIOFPANLQLPQGVLIK 54
Db TTSSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTE 677
55 SSGPLMIVSFOQT---VTRAETTSNITSRPAVPANPOTVKTICVPPNSSQLIKKAV- 109
Db SSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPV 737
110 -TPVKKLAIOGIVTT-----VPKPSVGS-----VANPTSVVYTPGKPLNF- 152
Db PTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTP 797
153 -----VTLKPSLSGASTPSNEPNLKAENSAAYOINTPTWLEWVKCKKNFLA 201
Db SSSSTTESSVAPVPTPSSSNITSSAPSSSTPSSSTTESSSVV---PTPSSSTTESSS--- 851
202 MLIKACGSGS-----PEKGONVKVLVQLLDAKTEAEFTFRKILYVELKSSPQHLV 254

```

Db      852  --APVSSSTESSVAPVPPPPSSSSNTSSADSSIDPFSSTESFSTGTVPSSSK----- 904
OY      255  PPLKSVVALROLPLPNSSOSFIQOCVOQTSMDVIATCTVVTSP---VTTTVSSSSOSE 311
Db      905  -----VYGSQT--ETSVSSTETTVYPTKTTTSVTTPSTTTITTTVCSTGCTN 949
OY      312  KSIIVSGATAPRTV--SVQTLNPLAGPVGAKAVVTLHSGPATAAGGTAGTG--LLQT 367
Db      950  SAGETTSQCSPTVTVTPPTTTTTSVTTSSTTTTITTTVCSTGCTNSAGETTSQCSPTKITY 1009
OY      368  SKPLTYSVANT-----VTVVSLQPEKPYVSGAVLISLPAV--FGEISGAIC----- 414
Db      1010  TVPCSTPSESEASASESTSPPTTPVTVVSTVTVTTERSSTKPGELITTFEVRKNIPITY 1069
OY      415  -----LPSVKPVSFCDMDHICKPVIGFPVOIKLAQPGVLSQAPGIPYSSSKQLPSLF 468
Db      1070  LTTIAPFTSVTIVTF-----TPTTTTT---VCS-----TGINS----- 1101
OY      469  HVVQPPSGGNEKQVTTTISHSSTLTIOKCGQKTMPTVNTIPTS 510
Db      1102  -AGETTSQCSPTVTVTPVPCSTGT---GEXTTEATLTIVTIA 1138

RESULT 8
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
C:Geisels, C.; Gattlung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34434
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GET>
A:Cross-references: EMBL:U08046; PIDN:NA070890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
A:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20

Query Match          7.7%  Score 212.5;  DB 2;  Length 2232;
Best Local Similarity 21.9%  Pred. No. 0.00039;
Matches 138;  Conservative 73;  Mismatches 247;  Indels 173;  Gaps 24;

OY      26  QIVAAKAPNTTITQIPANLQLPQGVTLIKNSGPLMLVPOQVTRAETTSNITSRAVP 85
Db      801  QSVSTNSPGSYTR-----PSTVSGSTSGSTVYVGTSEASTSGSSVASSPADSTS 852
OY      86  ANPQTWICIVPNSSSOLIKKVAATPVKKLAQIGTVTVTPKPSVQSVAAVPTSVTVT 145
Db      853  QNPN-----PSTSS-----GSSMTQSPYPS--QSTSPVSSSTTPS 886
OY      146  PKPLNTVYTLKPS-----SLGASSTPSN-----EPNL 173
Db      887  PESPPTTITLSTSPSPSGSTTIGSTQGSTPGISTSEEMTSQGSTQTPGSGYSTVTPST 946
OY      174  KAENAAA--VOINLS-----PTMLENVKCKKNFLAMLKLCAGS--OSPEMGON 219
Db      947  VSDSTSSSTVYVGTSEBSSSPIRSTSONTPST-----SGSSMSTQTPSSOS 996
OY      220  VKKIVQLLDAKIEEFTRLRYELKSSPOPH-----PPFL 257
Db      997  TSP-VESSGATSSSGSGPTLTLSISPPSPSSSTIGSGSGSTSPVVSITISQSTERTGS 1055
OY      258  KKSVALAQLL-----PN-SQSFIOOCVOQTSMDVIA 289
Db      1056  TGSVTKRPSVGSASGSGSTATMGSTEASSTGSGSTSPNSQSTSPSTSGATSSPSSSG 1115
OY      290  TCTTVVTPSPVTVTVSSSGSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAVVTLHSV 349

```


Db 926 TSAPASSSTSATVTVSSS-----QSPTSSPAQSSSTPSSSTVVOSSSFQ 975
Qy 498 --QKTMV--NTIIPQSOPPASI-----LKQITLPG--NKILSLQASSTOKMRI 541
Db 976 SPOSTIGSSTVTPSTQAPSSSTGGPTTQICPNQOTVFKGQGVYIEMLPASTQONAI 1035

Qy 542 K---ENV 545
Db 1036 NAFENV 1042

RESULT 11

A47282
calcium-binding protein calphotin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A47282
R:Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993
A:Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
A:Reference number: A47282; PMID:93165729; PMID:8094559
A:Accession: A47282
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-865 <MAR>
A:Cross-references: GB:L02111; MID:g157031; PIDN:AAA28405.1; PID:g157032
A:Experimental source: photoreceptor cells
A:Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBI:P:124956)
C:Genetics:
A:Gene: FlyBase:Cpn
A:Cross-references: FlyBase:FBgn010218
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: calcium binding

Query Match 7.5%; Score 206.5; DB 2; Length 865;
Best Local Similarity 21.7%; Pred. No. 0.00025;
Matches 122; Conservative 82; Mismatches 242; Indels 115; Gaps 22;

Qy 1 GTLVTKV-APVSAP-----PKVSSGRLPAP---QIVAVKAPNTTIOFPANIQ 45
Db 4 GTISPSAPVAPVYTSAAAPVQVSPAAVAPAPAPVAVTAPPTIASVQAPATVT 63
Qy 46 LPPTVLIKNSGRLVSPQOTVTRAETTSNITSRPAPAPQTVKICVTPNS-SSQLI 104
Db 64 IPAPAPIAASVTPVAVSAP--PVVAPTP-----PAASVSTPVAQIPVAVSAPVA 115
Qy 105 KKVAVTTPVKL-AGITVTVTVKPSVSVANPTSVT-VTGRKPLNTYTTLKPSLIG 162
Db 116 PPAVATPTPVQIPVAPVATPTPVAASAPTPAVTPEVISPVIASPPVAPANTTVPAAP 175
Qy 163 ASSTPSNEPILKAENSAAYOINLSPMLE-----NYKCKKNFLAMLIKACS 209
Db 176 VAAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 227
Qy 210 GSGSPKGVNVKLVLEQLDAKEEFTKRLYELKSSPOPHLVPLKKSVALROLPL 269
Db 228 -----PEVSVATKPLAAEPVAVPAPTEPVPVAPAAASPHVAVPAVERVA--P 278
Qy 270 NSQSFIOCCVQOQSSDMVATCTTTTSPVYTTVSSSQSEKSIIVSGATAPRTVSQY 329
Db 279 VVAS-----TEPPVAAATLTTPAPT-PALAPVAVESQVAVANTVATPTP-APEPET 328
Qy 330 LNPAGPVGAKAGVVLTHSVGPTAATGCTTAGTGLTQSKPLVSNVNTVTVLSQPEKP 389
Db 329 IAP-----PVVAETPEVASVAVAEETTPVVPVA-----AESIPAP 364
Qy 390 VSGTAATVLSLPAVTFETSGAALCLPVSVPVVS-----FCMDHICKPVYGT 436
Db 365 VAAVTPVPAVL-AVTDVDDVASAV--DELPPVIAVSPVPAVPAVPAVPAVPAVPAV 421
Qy 437 PVQIKLAPGPFVLSQAPGIPFGSSSKQLFLSHVVOQPSGGNEKQVTTTISHSLTLLTK 496

Db 422 PVPAVVAEETPETPAPASAPVTLALDIPVAVPAVPAVPAVPAVPAVPAVPAV 473
Qy 497 GQKTMVNTIIPQSOPPASI 517
Db 474 -VSTPPTTASVPETTAPPAV 493

RESULT 12

T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34433
R:Giesel, C.; Gattung, S.
submitted to the EMBL data library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34433
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1032 <GEI>
A:Cross-references: EMBL:U00846; PIDN:AAC70889.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: x
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2;
Query Match 7.4%; Score 205; DB 2; Length 1032;
Best Local Similarity 22.7%; Pred. No. 0.00038;
Matches 133; Conservative 74; Mismatches 227; Indels 152; Gaps 24;

Qy 2 TLVTKVAPVSPRVSSGRLPAP---QIVAVKAPNTTIOFPANIQLPCTVLIKNSG 58
Db 246 SITSALPT-ASSSASSPASAASSTTPVLISSSTIOSSGTFPSSVASSPSTV--GSTSG 302
Qy 59 PLMLVSPQOTVTRAETT-SNITSRPAPAPQTVKICVTPSSSQLIKKAVTPEVKLIAQ 117
Db 303 AASSSVATVSTINGSTGTTT---PVGSSSTIGSSSTPASAASSSSSGTMSIS-----GS 354
Qy 118 IGTTVTVTVKPPSS-----VQSVAVPTSVYVTPGKPLNTYTTLKPSLIGASPTBNE 170
Db 355 TGSF-VTVVPGSSSTFASSTPIASSSSGSTVTAFGS-----SSTVYGSSTPS-- 401
Qy 171 PNLKAENSAAYOINLSPMLENVKCKNFLMLIKACSGSGSPKGVNVKLVLEQLDA 230
Db 402 -----ASSSSSGTMTNSGTSSTVYA 424
Qy 231 KIEAEFTKRLYELKSSPOPHLVPLKKSVALROLPLNSQSFIOCCVQOQSSDMVAT 290
Db 425 PVSSSTFG-----SSP-----IASSSSGSTVTVVSGSSSTVYGS 460
Qy 291 CTTTIVTSPVYTTVSSSQSEKSIIVSGATAPRTVSQTLNP-----LAGVGAAG 342
Db 461 TPSASSSAGTASTISGTSATIVPSSSSVSSSTQASAPSSPGTMTVSGTGTSTVT 520
Qy 343 VTLHVSQPT-----AATGCTTAGTGLTQSKPLVSNVNTVTVLSQPEKPYV 391
Db 521 VVPGSSSTPAPSSSPNPSASPASTGSTITTTISG-----SSIIIVTVSGS-----TV 566
Qy 392 SGAATVLSLPAVTFETSGAALCLP-SVKPVVSFCMDHICKPVYGT--TPVQIKLAPGPV 448
Db 567 SGTSTGTSSTIASSTATPFGSSSTVPSSSSPQS---SGSPAPNPGTTPSTQSSQSPSPS 623
Qy 449 LSPAGIPTGSSSKQLFLSHVVOQPSGGNEKQVTTTISHSLTIT--OKGQKTMVNTI 506
Db 624 MNPSSSTPTGSSOGTITPEGSTASSPTG-----STGSTFVAVAEVTSQSTVPSGSS 674
Qy 507 I---PTSGPPASILKQITLPGKNKILSLQAS--TOKRKIKENTTS 547
Db 675 LGTOSTNSPSPSSLSPT-SGMSTLTSEPSSTQSGAOSTLTT 719

RESULT 13

DNA-binding protein R kappa B - human
C:Species: Homo sapiens (man)
C:Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 01-Dec-2000
C:Accession: S52863; A45580
R:Nieters, A.; Bouwmeester, T.; Scheldereit, C.
submitted to the EMBL Data Library, August 1994
A:Reference number: S52863
A:Accession: S52863
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1324 <NIE>
A:Cross-references: EMBL:X80878; NID:6695578; PID:6695579
R:Adams, B.S.; Leung, K.Y.; Hanley, E.W.; Nebel, G.J.
New Biol. 3, 1063-1073, 1991
A:Title: Cloning of R kappa B, a novel DNA-binding protein that recognizes the interleukin
A:Reference number: A45580; MUID:92135142; PMID:1777480
A:Accession: A45580
A:Molecule type: mRNA
A:Residues: 313-1292, 'ISNRQLLAP', 1304, 'QSPILL', <ADA>
A:Cross-references: GB:008191; GB:S79520; NID:9476273; PIDN:AAAL1871.1; PID:9476274
A:Note: sequence extracted from NCBI backbone (NCBIN:79520)
C:Keywords: DNA binding; transcription factor

RESULT 14

nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: J30826
R:Yotov, W. V.; St-Arnaud, R
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8658236
A:Accession: J30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: EMBL:U048363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Intons: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding ac
:Keywords: alternative splicing; DNA binding; transcription factor

RESULT 15

probable Myb-family transcription factor F16M22.4 [imported] - *Arabidopsis thaliana*
C;Species: *Arabidopsis thaliana* (mouse-ear cress)

...

...

...

...

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 10.6381 Seconds
(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552
Perfect score: 2758
Sequence: 1 GTLVTKVAPVAPKVVSSGP.....ASPTOKRIKENVTSCEFRDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	2758	100.0	801	T2DT_HUMAN	Q92750 homo sapien
2	547	19.8	1083	T2D3_HUMAN	Q00268 homo sapien
3	256	9.3	921	T2D3_DROME	P47825 drosophila
4	223.5	8.1	2035	HFC1_HUMAN	P51610 homo sapien
5	223	8.1	2090	HFC1_MESAU	P51611 mesocricetu
6	220	8.0	1367	AMYL_YEAST	P08640 saccharomyc
7	219	7.9	5179	MUC2_HUMAN	Q02817 homo sapien
8	206.5	7.5	865	CPN_DROME	Q02910 drosophila
9	197.5	7.2	5376	ZAN_MOUSE	O08799 mus musculu
10	195	7.1	1161	DAN4_YEAST	P47179 saccharomyc
11	195	7.1	1322	YAG3_YEAST	P39712 saccharomyc
12	194.5	7.1	2700	ZAN_HUMAN	O94933 homo sapien
13	193.5	7.0	670	VG50_HSV1	Q00130 ictaluriid h
14	189.5	6.9	797	VGIX_HSV1	P28968 equine herp
15	184.5	6.7	881	JUH8_YEAST	P47033 saccharomyc
16	181.5	6.6	662	MUC1_XENLA	Q05049 xenopus lae
17	181.5	6.6	1199	N121_RAT	P52591 rattus norv
18	181	6.6	1609	FTG2_YEAST	P25653 saccharomyc
19	180.5	6.5	725	AGAL_YEAST	P32323 saccharomyc
20	180.5	6.5	3178	YS89_CABEL	Q09664 caenorhabdi
21	179.5	6.5	1537	FL01_YEAST	P32768 saccharomyc
22	178.5	6.5	1260	ALST1_CANAL	P46590 candida alb
23	175.5	6.4	2090	N124_HUMAN	P35658 homo sapien
24	174	6.3	1075	FLOS_YEAST	P38894 saccharomyc
25	172.5	6.3	1858	P3K2_DICDI	P54674 dictyostell
26	172	6.2	886	VG3_EBYA8	Q07284 epstein-bar
27	172	6.2	1169	YK82_YEAST	P36170 saccharomyc
28	169.5	6.1	745	OCT1_PIG	Q29076 sus scrofa
29	169	6.1	1229	N121_HUMAN	O94263 homo sapien
30	168	6.1	1306	MSB2_YEAST	P32334 saccharomyc
31	167.5	6.1	1140	YMG6_YEAST	Q04883 saccharomyc
32	166.5	6.0	743	OCT1_HUMAN	P14859 homo sapien
33	166.5	6.0	3866	HRX_MOUSE	P55200 mus musculu

34	166	6.0	3726	1	TRX_DROME	P20659 drosophila
35	165	6.0	606	1	SP2_HUMAN	Q02086 homo sapien
36	164.5	6.0	1365	1	SUZ2_DROME	P25172 drosophila
37	163	5.9	860	1	CH12_COCIM	P54197 coccidioid
38	162	5.9	3969	1	HRX_HUMAN	Q03164 homo sapien
39	161	5.8	1119	1	AL53_CANAL	O74623 candida alb
40	161	5.8	1509	1	GSRI_HUMAN	O94264 mus musculu
41	160.5	5.8	630	1	MUC1_MOUSE	Q02496 mus musculu
42	159	5.8	676	1	MUC1_MESAU	O60528 mesocricetu
43	157.5	5.7	907	1	VG3_EBY	P03200 epstein-bar
44	153	5.5	739	1	OCT1_CHICK	P15143 gallus gall
45	152	5.5	604	1	MTG8_HUMAN	Q06455 homo sapien

ALIGNMENTS

```

RESULT 1
ID T2DT_HUMAN STANDARD; PRT; 801 AA.
AC Q92750:
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor TFIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=97011146; PubMed=8858156;
RA Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIID subunit related to
   hTAFII130."
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
CC A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
CC MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
CC PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
CC HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y09321; CA70499.1; -
CC Genew; HGNC:11538; TAF4B.
CC MIM; 601689; -
CC DR InterPro: IPR003894; TAF_hom.
CC SMART: SM00549; TAFH; 1.
CC KW Transcription regulation; Nuclear protein.
CC FT NON_TER
CC SQ SEQUENCE 801 AA; 85658 MW; D1284932FEA9CD2 CRC64;

Query Match 100.0%; Score 2758; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 8.3e-147;
Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GTLVTKVAPVAPKVVSSGRLPAPOIVAVKAPMTTITQFPANLQLPQGVLIKNSGSL 60
|||||

```

DB 1 GRLVTKVAVSAPKRVSSGRLPAPQIVAVKAPNTTIOFANLQLPSTGLVLIKSNGPL 60
 OY 61 MLVSPQOQYTRARETTSNITSRAVPANPOTVATCTVPSNSSQILKKVATPVYKKAQIGT 120
 DB 61 MLVSPQOQYTRARETTSNITSRAVPANPOTVATCTVPSNSSQILKKVATPVYKKAQIGT 120
 OY 121 TVTIVTPKRSVQSAVAVPTSVTPGKPLNTVTLTKPSLSGASSTPSENPILKENSNA 180
 DB 121 TVTIVTPKRSVQSAVAVPTSVTPGKPLNTVTLTKPSLSGASSTPSENPILKENSNA 180
 OY 181 VOINLSPTMLENVKKCKNEFLAMLIKACSGSOSPEMGQNVKRLVQLDLDAKTEAEETFRK 240
 DB 181 VOINLSPTMLENVKKCKNEFLAMLIKACSGSOSPEMGQNVKRLVQLDLDAKTEAEETFRK 240
 OY 241 LVELKSSQPHLVPFLKSSVVALQQLPNSOSFIOQCQOQSSDPVATCTTIVTTSV 300
 DB 241 LVELKSSQPHLVPFLKSSVVALQQLPNSOSFIOQCQOQSSDPVATCTTIVTTSV 300
 OY 301 VTTVSSSQSEKSIIVSGATAPRTVSVOTLNPFLAGVGAAGVTLHSVGTATGTTA 360
 DB 301 VTTVSSSQSEKSIIVSGATAPRTVSVOTLNPFLAGVGAAGVTLHSVGTATGTTA 360
 OY 361 GTGLIQTSPKPLVTSVANTVTVSLQPEKPVSGTAVTSLSPAVTEGETSGAAILCLPSVXP 420
 DB 361 GTGLIQTSPKPLVTSVANTVTVSLQPEKPVSGTAVTSLSPAVTEGETSGAAILCLPSVXP 420
 OY 421 VVSFPCMDHCKRPGVIGTPVQIKLAOPGVLISOPAGIPTGSSSQQLPSLFHVVOQPSGNGK 480
 DB 421 VVSFPCMDHCKRPGVIGTPVQIKLAOPGVLISOPAGIPTGSSSQQLPSLFHVVOQPSGNGK 480
 OY 481 QVTTSISSTLTIOKCGQKTMPTNTIIPTSOPPPASILKQITLPGKRIISLOASPTOKNR 540
 DB 481 QVTTSISSTLTIOKCGQKTMPTNTIIPTSOPPPASILKQITLPGKRIISLOASPTOKNR 540
 OY 541 IKENVTSCFRDE 552
 DB 541 IKENVTSCFRDE 552

RESULT 2
 ID T2D3_HUMAN STANDARD; PRT: 1083 AA.
 AC 000268; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription Initiation factor TFIIID 135 kDa subunit (TAFII-135)
 DE (TAFII135) (TAFII-130) (TAFII130).
 GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97336072; PubMed=9192867;
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Buttrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.D., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Graffam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.D.,
 RA Humble E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levasaibo M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.E.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurray A.A.,
 RA Milne S.A., Misty D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prichaltingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Selha H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann K.C., Sycamore N., Taylor R., Taylor R., Thomas D.W., Thorpe A.,
 RA Tracey A.L., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilting L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RL "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN (3)
 RP SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=97098442; PubMed=8942982;
 RA Tanase N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
 RT "Molecular cloning and analysis of two subunits of the human TFIIID
 RL complex: hTAFII130 and hTAFII100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC - FUNCTION: MAKES PART OF TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT
 CC PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS
 CC ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION
 CC BY THE AP-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC - SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFS).
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: Y11354; CAA72189.1; -
 CC EMBL: A1137077; CAC36006.1; -
 CC EMBL: A1109911; CAC22312.2; -
 CC EMBL: U75308; AAC50901.1; -
 CC TRANSFAC: T02328; -
 CC DR GeneW: HGNC:11537; TAF4.
 CC DR MIM: 601796; -
 CC DR InterPro: IPR003894; TAF_hom.
 CC DR SMART: SM00549; TAFH; 1.
 CC DR Transcription regulation: Nuclear protein.
 CC KW DOMAIN 39 42
 CC FT DOMAIN 52 57
 CC FT DOMAIN 98 101
 CC FT DOMAIN 142 148
 CC FT DOMAIN 268 275
 CC FT DOMAIN 331 337
 CC FT DOMAIN 680 683
 CC FT DOMAIN 808 813
 CC FT DOMAIN 828 831
 CC FT DOMAIN 105 117
 CC FT
 CC FT CONFLICT 136 136
 CC FT CONFLICT 185 185
 CC FT CONFLICT 233 264
 CC FT CONFLICT 293 293
 CC FT
 CC FT SEQUENCE 1083 AA; 109943 MW; A645382575A0752 CRC64;
 SQ
 Query Match 19.8%; Score 547; DB 1; Length 1083;
 Best Local Similarity 30.2%; Pred. No. 2,4e-23;
 Matches 173; Conservative 67; Mismatches 144; Indels 188; Gaps 21;


```

QY 13 PKVSSG-----PRLPAPQIYAVKAPNTTTTIOFPANLQLPQCTVLIKNSGFLM 61
DB 414 PTAATSGIRATLPPTVLAPRPOP-----PQNPPTNIO---NFQLPQGMVLVRESCOLL 464
QY 62 LVSPQOQVTR-----AETSNITSBPVAVNPQFVKICVNPSSSOLIKKVAATPVKKL 115
DB 465 MI-POCLADQAOAOAHQAPQPTTMAPPRATPTAPPOVISTVQADGTPILAN-QVTP----- 518
QY 116 AQIQVTVTVTVTPKPSVQSAVPTSVTVTPGKPLNTVT--TLKPSISGASS-----TPS 168
DB 519 -----TTIIKQV---SQAQTTVPQSATIQRSQVQPOLVIGSAQATFASLGTFANQOTGTPQ 571
QY 169 NE-PNLKAENSAVOILSLPTMLENKCKNFLMLIKLACSGSQSPMGONVKKVEQL 227
DB 572 RTVPQATTTSSAATE-----TMENVKCKNKFSLTLIKILASSCKOSTETANVKEVLQNL 625
QY 228 LDKATIEEETRLKLYELKSSPOPHLPFLKSSVVALROLPLNQSOSTLOQVOO-----TS 283
DB 626 LDGKIEADETTRILKRLNSSPOPTLVPEFLKSLPALRQLPDSAAFIQSQOQPPPTTS 685
QY 284 SDMVIACTTTVTTPVTVTVSSSSOSEKSIIVSGATAPRTVSQVOTLNPLAGPVKAKGV 343
DB 686 Q-----ATTALTAVLVSSSVQRTAKTAATVTSALQPPVLSL----- 722
QY 344 VTLHSVGPATAGTGTAAGTGLQTSKPLVTSVANTTVTSLOPEKPVVSGTAVTSLPVA 403
DB 723 -----TQPTQVGVKQOGPPPLVIO-----QPKRP-----GALIRPPQV 756
QY 404 TFGFTSGAATCLPSPKRVVSEFCMDHICKPVIGPPVQIKLQAPGVLSQAPAGIPGSSSKQ 463
DB 757 TLTQVT-----PVALRQDP-NRIMLTTPQOIQOL----- 783
QY 464 LFSLFHVQOPSGNGEKQVTTSHSSTLTIOCKGQKQKMPVTTIPTSQFPASILIKQITL 523
DB 784 ----- 798
QY 524 PGKIL---SLOASPTOKNRIKENVTSCEPDE 552
DB 798 PGTKALSAVSAOAAAQKKNLKEPGGSGFRDD 829

```

RESULT 3

T2D3_DROME STANDARD: PRT: 921 AA.

AC P47825; P49845; Q9YUY7;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transcription initiation factor TFIID 110 kDa subunit (P110)

DE (TAF11-110) (110 kDa TBP-associated factor).

GN TAF110 OR CG5444.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephyridioidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 519-540; 597-616 AND 857-874.

RC TISSUE=Embryo;

RA MEDLINE=93145326; PubMed=7678780;

RA Hoey T., Weinzierl R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,

RA Tjian R.;

RT Molecular cloning and functional analysis of Drosophila TAF110

RT reveal properties expected of coactivators.;

RL Cell 72:247-260(1993).

RN [2]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 398-406; 520-540 AND 860-877.

RC TISSUE=Embryo;

RA MEDLINE=93317591; PubMed=8327460;

RA Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.;

RT The Drosophila 110-kDa transcription factor TFIID subunit directly

RT interacts with the N-terminal region of the 230-kDa subunit.;

RL Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Holt C., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolashkov S.,
RA Borokova D., Botchan M.A., Bouck J., Brockstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Styksas R., Tector C., Turner R., Venter R., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster.;"
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
CC CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
CC AND REPRESSORS. MAY FUNCTION AS A COACTIVATOR BY SERVING AS A SITE
CC OF PROTEIN-PROTEIN CONTACT BETWEEN ACTIVATORS LIKE SPI (OR BTI)
CC AND TFIID COMPLEX.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).


CC EMBL: L06861; -; NOT ANNOTATED_CDS.



DR EMBL: S63550; AAB27433.1; -



DR EMBL: AE003528; AAF49536.1; -



DR TRANSFAC: T02121; -



DR FlyBase: FBgn010280; TAF110.



DR InterPro: IPR003894; TAF_hom.



DR SMART: SM00549; TAFH; 1.



KW Transcription regulation; Nuclear protein.



FT DOMAIN 66 82



FT DOMAIN 108 111 POLY-GLN.



FT DOMAIN 259 265 POLY-GLN.



FT SEQUENCE 921 AA; 99338 MW; 27E6852859872767 CRC64;


```

Query Match 9.3%; Score 256; DB 1; Length 921;
 Best Local Similarity 22.9%; Pred. No. 3.3e-07;
 Matches 135; Conservative 86; Mismatches 193; Indels 176; Gaps 24;

25 PQIVAVKAPNTTITQPPANLQPPGVILKSNSSGPMML-----VSPQGVTRAEVTS 76
 133 PPSSTITLSTLNLGQPPA-----LVKTONGFOLLRGVTTGPTVQTITNNSNS 184
 77 NTSRAVAVANPQTVKICVVPNSSQ-----LKKAVVPVKKLAQIGTVVTVTP 127
 185 NNTSTNNHPTTQ-IRLQVVPAAASMTNTTATNIIIVNSVASSGVANSSQPPHQLQMAQ 243
 128 KSSSVAVAPTSVTVTGGKPLNTVTLKPSISLASSPSPNPRLKAENSAVQINLSP 187
 244 ARLPQITQITQIPAAQSOQOVNANVSSAGGTAVASSTA-----ATT 287
 188 TMLENVK-KCKNFLMLIKLACSGSSPENGQVKKLVOLDLAKTEAEERKLVYELK 246
 288 TQGNTRKCKRKLMLIEL--STRBKPYEKVNTLLQELVANAVEPEEFCORLERLN 345
 247 SSPQHLVFFLKKSVVALQQL-----LPNSQSFID--- 276
 346 ASPQCLIFLKKSLPLRLQALYKELVLEGIRPPQHVLAGLSOQLPKIQAIPIG 405
 277 ---QCVQSTSDMTATCTTIVTSPVTTTSSQSEKSIIVSGATAPRTVS---VOT 329
 406 PSETTITIGOTQVMT--TPRALCTPRPTTIGHTTISKOPR---IRLPTRALVNTGIGRT 460
 330 LNPACPVGAKAGVTLHVSVPRTAAGTGTAGTLQTSKPLVTSVANTVTVSLOPEK 389
 461 QIP-SIQVGGANIVQIR--GPOHQLQRTGSGVQIRATRP-----PNSVPTAN----- 506
 390 VVSGTAVTSLPVTGEGISGAICLPVSKPVVSPFMDHICRPVIGTPVQIKIQAQEPVL 449
 507 -----KLTAVKVGGQIKAI-TPSLHP-----PSL 530
 450 SOPAGIPTSSSKQSLFLFHVYQQPSGNEKQVTLTSHSSTLTQKCGKTMPTNTIIP 509
 531 AMSGPPPTPLTSLVS-----TLNAST-----TTLPIS-LPT 564
 510 SOPPAST---LKQITLPGN-----KILSLQA---SPTOKRIKENVT 546
 565 VHLPEALRAREQMONSLNHSNHDAKIVEIKAPSLHPHMERINASTL 614

RESULT 4
 HFC1_HUMAN STANDARD; PRT; 2035 AA.
 AC P51610;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF) (CFP).
 GN HFC1 OR HCF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND PARTIAL SEQUENCE.
 RX MEDLINE=93327419; PubMed=8392914;
 RA Wilson A.C., Iamarco K., Peterson M.G., Herr W.;
 RT "The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein.";
 RL Cell 74:115-125(1993).
 RC [2]
 RP SEQUENCE OF 65-2035 FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Fetal brain;
 RA MEDLINE=95130107; PubMed=7829097;
 RA Fratini A., Faranda S., Redolfi E., Zucchi I., Villa A., Patrosso M.C., Strina D., Susani L., Vezoni P.;
 RT "Genomic organization of the human VP16 accessory protein, a

housekeeping gene (HFC1) mapping to Xq28.";
 RT Genomics 23:30-35(1994).
 RN [3]
 RP PROCESSING.
 RX MEDLINE=96033796; PubMed=7590226;
 RA Wilson A.C., Peterson M.G., Herr W.;
 RT "The HCF repeat is an unusual proteolytic cleavage signal.";
 RL Gene Dev. 9:2445-2458(1995).
 CC -1- FUNCTION: UPON LYRIC INFECTION OF PERMISSIVE CELLS, THE HSV TRANSLACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.
 CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N- AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBERT NONCOVALENTLY, ASSOCIATED.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A SHORT FORM, MAY BE PRODUCED BY ALTERNATIVE SPLICING. IN THE SHORTER FORM THE N- AND THE C-TERMINAL FRAGMENTS FAIL TO ASSOCIATE.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL TISSUES AND THE ADULT KIDNEY. PRESENT IN ALL TISSUES TESTED.
 CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE SIGNAL.
 CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS AT A DEFINED SITE, PCE/THER, WITHIN THE HCF REPEAT.
 CC -1- PTM: GLYCOSYLATED. CONTAINS O-LINKED N-ACETYLGLUCOSAMINE RESIDUES.
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL; L20010; -; NOT_ANNOTATED_CDS.
 DR EMBL; X79198; CAA55790.1; ALT_INIT.
 DR Genew; HGNC:4839; HFC1.
 DR MIM; 300019; -
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF01344; Kelch; 5.
 DR SMART; SM00060; FN3; 1.
 KM Nuclear protein; Repeat; Alternative splicing; Glycoprotein.
 FT REPEAT 44 89
 FT REPEAT 93 140
 FT REPEAT 148 194
 FT REPEAT 217 265
 FT REPEAT 266 313
 FT DOMAIN 1010 1439
 FT REPEAT 1010 1035
 FT REPEAT 1072 1097
 FT REPEAT 1101 1126
 FT REPEAT 1158 1183
 FT REPEAT 1286 1311
 FT REPEAT 1314 1339
 FT REPEAT 1349 1374
 FT REPEAT 1414 1439
 FT REPEAT 382 450
 FT VARSPLIC 564 564
 FT CONFLICT 603 603
 FT CONFLICT 1164 1164
 FT CONFLICT 1873 1873
 SQ SEQUENCE 2035 AA; 208841 MW; 99207FBB875204C0 CRC64;

Query Match 8.1%; Score 223.5; DB 1; Length 2035;
 Best Local Similarity 22.9%; Pred. No. 5.7e-05;

Matches 147; Conservative 79; Mismatches 217; Indels 199; Gaps 33;

QY 2 TLVTKAVAPSPKSSGRPLP-----APQIVAKAPNTT-----IQ 39

Db 446 TLVTPQAP--APPTTTIQLPVPSSISVPARTQGVPAVLKVPQATTTPLVTM 503

QY 40 FPN-----LQDPGVLI---KNSGPMLVSPQOT---VTRAETTSNI--TS 80

Db 504 RPASQAKAPVTVSLPAGVPMVPTQSAQGVTVIGSSPQSGMAALAAAAATQKIPSS 563

QY 81 RP--AVPN-----PQVKKICV--NSSQILKKVAVTPVKLAQ 117

Db 564 RPVYLSVPAGTIVTKMAVTPGTTLPATVAVASSPVMVSNPATRMLKTA-----AQ 616

QY 118 IGTVV-----VTVPRKSSVQSAVPTSVTVTPGKPLMTVTLK--PSSL--GASS 165

Db 617 VGRSVSATMTSTRTPLITVHKSGTV--TVAGQAVTVTVGVTMTITLVKSPISVPGSA 675

QY 166 TFSNEPML-----KAENSAVQVINSPTMLENVKCKKFL--AMLIKLACSGSQSP-- 214

Db 676 LISMLCKVMSVQTKPVQTSVAVTGOASTGCVTQIITQKGPLPAGTIIKLVTSSADCKPPTI 735

QY 215 -----EMGQVKKLVQDLDAKIEEFTRKLYELKSPQPHLYPPLKKSVALRQLP 269

Db 736 ITTQASGAGTKPTILCI-----SSVSPSTT---KPGTTTIITIP 773

QY 270 NSQSFIOQCVQOTSSDVIACTTPTVT-----SPVVT-----TVSSSQSEKSI 314

Db 774 MSALITQAGATGVTVSSPGISPTIITTKVMTSGTGAAPAKITAVPKATIGHQGGQVQV 833

QY 315 IVGSA-----TAPRT-----VSQTLNPLAGPVGAK--AGVYTLHSVGPATAG 356

Db 834 VLKAGAPQCPGTLITRTVPMGCVRLVTPVTVSAVKPAAVTLVKGTTGVTLCGTAVGTST- 892

QY 357 GTTAGTLLQTSKPLVSA--NTVTVSLQPEKPVYSGRAVTLSLAVTGESGAIC 414

Db 893 -SLAGAGHSTSLAPLITTLGITATLSSO---VINPALTIVSAAGTILTAAGLTPP 947

QY 415 LPSVKPVVSPFCMDHICKPVIGTVPQIKL--AOPGVLSQPA--GIPTGSSSKQLFSLFHVQ 472

Db 948 TIMQAPV-----SPTQVTLITAPSGVEAQPVDLPVS-----ILA 983

QY 473 QPSGNEKQVTTISHSTLTITQCKGQKTPMVPNTIIPSPFP 514

Db 984 SP-----TTEDPRTAVTIVASGQGVQPGTIVLCSNP 1017

RESULT 5

HFC1_MESAU STANDARD; PRT; 2090 AA.

ID HFC1_MESAU STANDARD; PRT; 2090 AA.

AC P51611;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFC1) (VCAF) (CEF).

GN HCF1.

OS Mesocricetus auratus (Golden hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.

OC Mesocricetus.

OX NCB1_TaxID=10036;

RN NCB1_TaxID=10036;

RP SEQUENCE FROM N.A.

RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY SIMILARITY).

CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM

CC 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N- AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE SIGNAL (BY SIMILARITY).

CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS AT A DEFINED SITE, PCEC/THER, WITHIN THE HCF REPEAT (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; D45419; BAA08258.1; -.

DR InterPro: IPR003961; FN_III.

DR InterPro: IPR001798; Kelch.

DR Pfam; PF00041; fn3; 2.

DR Pfam; PF01344; Kelch; 5.

DR SMART; SM00060; FN3; 2.

KW Nuclear protein; Repeat.

FT REPEAT 44 89

FT REPEAT 93 140

FT REPEAT 148 194

FT REPEAT 217 265

FT REPEAT 266 313

FT DOMAIN 1010 1448

FT REPEAT 1010 1035

FT REPEAT 1072 1097

FT REPEAT 1101 1126

FT REPEAT 1157 1182

FT REPEAT 1295 1320

FT REPEAT 1323 1348

FT REPEAT 1358 1383

FT REPEAT 1423 1448

FT SEQUENCE 2090 AA; 214942 MW; E495EBB1F2385E17 CRC64;

Query Match 8.1%; Score 223; DB 1; Length 2090;

Best Local Similarity 22.0%; Pred. No. 6,2e-05;

Matches 130; Conservative 75; Mismatches 190; Indels 196; Gaps 27;

QY 10 VSAPPKVSQPRLPAPQIVAKAPNTTTIOPANLQLEPGTILKSNQPMLVSPQQT 69

Db 537 IGSNPOKSGMAALAAAAATQKIPSSA--PTVLSVPAGTIVKT-----VAVTPTGTI 588

QY 70 TRAFETSIITSRPVAPNPQ-----VAICVPPNSSQ---LIKVAVTPYKKL 115

Db 589 LPA--TVKAVASSPVMSVPATRLMLKTAAGVGTSSSAANTSTPTITVHKSGTVAAQ 646

QY 116 AQISTV-----TVPKP-----SSVQVAVPVS-----140

Db 647 AQVTVTVGVTKRTITLVKSPISVPGSALISNLKQVMSVQTPVQTSVAVTGOASTGCPV 706

QY 141 -----VTVTPGKPLNTVTL-----KPSISGASST-----PSNE 170

Db 707 TQIQTGKPLPAGTILKLVTSADGKPTIITTTQASGASKPTILIGISSVSPSTTKGT 766

QY 171 PNLAENSAVQVINSPTMLENVKCKKFL--IKLACSSQSPQSGMAALAAAAATQKIPSS 229

Db 767 TITTPMSALITQAGATGVSTPGIKSPITITTKVTSCTGAP-----AKITAV-- 818

QY 230 AKIAEEFTRL--VVELKSSP--OPHLVPELKKSVVALROLPSQSFIOQCVQOTSSDMV 287

Db 819 PKIKTGQCGCVTVVLKAGAGQG-----AIKRTV-----MSGVR 855

QY 288 IATCTTIVTSPVTVTVSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVYTLH 347

RA MEDLINE=94132002; PubMed=8300571.
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-W., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -1- PAM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC OF SLIKMORM HEMOCYTIN.
CC -1- SIMILARITY: CONTAINS 2 WMFC DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE (CTCK) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L21998; AAB95295.1; -
DR EMBL: M74027; AAB59875.1; -
DR EMBL: M94131; AAB59163.1; -
DR EMBL: M94132; AAB59164.1; -
DR Genew; HGNC:7512; MUC2.
DR MIM: 158370; -
DR InterPro: IPR000359; Cys_knot.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002400; GE_cysknot.
DR InterPro: IPR002919; TIL_Cysrich.
DR InterPro: IPR001007; WMF_C.
DR InterPro: IPR001846; WMF_D.
DR Pfam: PR00007; Cys_knot; 1.
DR Pfam: PR00093; VWC; 1.
DR Pfam: PR00094; VWD; 4.
DR Pfam: PR01826; TIL; 1.
DR PRINTS; PR00438; GFCYSKNOT.
DR SMART: SM00214; VMC; 2.
DR SMART: SM00216; VMC; 4.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; WMFC; 2.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 5179 MUCIN 2.
FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
FT REPEAT 1401 1416 1.
FT REPEAT 1417 1432 2.
FT REPEAT 1433 1448 3.
FT REPEAT 1449 1464 4.
FT REPEAT 1465 1471 5.
FT REPEAT 1472 1478 6.
FT REPEAT 1479 1494 7A.
FT REPEAT 1495 1517 7B.
FT REPEAT 1518 1533 8A.
FT REPEAT 1534 1556 8B.
FT REPEAT 1557 1572 9A.
FT REPEAT 1573 1596 9B.
FT REPEAT 1597 1612 10A.
FT REPEAT 1613 1635 10B.
FT REPEAT 1636 1651 11A.
FT REPEAT 1652 1675 11B.
FT REPEAT 1676 1683 12.
FT REPEAT 1684 1699 13.
FT REPEAT 1700 1715 14.
FT REPEAT 1716 1731 15.
FT REPEAT 1732 1747 16.
FT DOMAIN 4815 4886 WMFC 1.
FT DOMAIN 4924 4991 WMFC 2.
FT DISULFID 5075 5122 CTCK.
FT DISULFID 5089 5136 BY SIMILARITY.
FT DISULFID 5098 5152 BY SIMILARITY.
FT DISULFID 5102 5154 BY SIMILARITY.
FT DISULFID 5102 5159 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1351 1351 H -> L (IN REF. 3).
FT CONFLICT 1412 1412 T -> S (IN REF. 3).
FT CONFLICT 1449 1449 L -> P (IN REF. 3).
FT CONFLICT 1504 1504 M -> T (IN REF. 3).
FT CONFLICT 4192 4192 G -> S (IN REF. 2).
SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571EB9A5663 CRC64;

Query Match 7.9%; Score 219; DB 1; Length 5179;
Best Local Similarity 21.9%; Pred. No. 0.00031;
Matches 136; Conservative 63; Mismatches 227; Indels 194; Gaps 29;
QY 2 TLVTKVAVPSAP-----PKVSSGPRLLPAPQIVAVKADNTTIT-----QEPAN 43


```

FT CARBOHYD 3065 3065 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3144 3144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3172 3172 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3288 3288 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3292 3292 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3782 3782 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4005 4005 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4136 4136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4243 4243 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4254 4254 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4335 4335 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4376 4376 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4586 4586 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 5136 5136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 5252 5252 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 5376 AA; 579908 MW; 0E4ADB77DE2A2620 CRC64;

Query Match 7.2%; Score 197.5; DB 1; Length 5376;
Best Local Similarity 23.2%; Pred. No. 0.0052;
Matches 146; Conservative 79; Mismatches 234; Indels 171; Gaps 33;

QY 8 APVSAB--PKVSGPRLPAPQIVAVKAPNTTITQFPA---NLQLPFGVLLIKNSGGLM 61
DB 551 SPLPPTGPESEVPLPMQPTSPKATVTITETPTTEATITPTTETTVTE---VI 607
QY 62 LVSPOQTVAETTSNITSRAVPANQYKICTVPSNSQLIKVAVPYKLAQIGTT 121
DB 608 MSPKETSIPPEVT-----IPTEVTV-----SPEELISPEVTPVDVTAAYV 652
QY 122 VWTTV-PKPSVY-QSVAVPTSVYTPGKRLNTVTLKPSST-GASTPENENLKA-E 177
DB 653 EATNASPEETSVPEVTITLEVTVSPEE--TTPPEVPIVLIDAPFAPEGETTLTEVP 710
QY 178 SAAVOINLSPTMLNKKCNFLAMLIKLAGSGSPSEMQNKVKEQLDLAK---IEA 234
DB 711 TVPTEVTVGHTEVTV-----SPE---EHSVPEEISTEVTVSP 748
QY 235 EETTRK-----LYVELKSSQPHL-----VPFLKSVVALROLPLPSOSFTQOCVOOTS- 283
DB 749 EETTVPEVPIVLEATASPTGEITLYTEVPVPEVTVGHTEVTVSP-----EETSV 802
QY 284 -SDMVIATCTTV-----TTSPTVTTVSSSGSEKSIIVSGATAPRVYQVOTLPLAGPVG 338
DB 803 PHEETISTEVTYTSPEETITLPTVEPTVSTEVTVS-----PEETSVPEETI----- 849
QY 339 AKAAGVTLHSVGPPTAATGGTTAGTGLL-----QTSKPLVTSVANTVTVS-----IQPEK 388
DB 850 ----LTTLYTEVPTVPEVTVGHTEVTVNSPEETSVPEETISTEVTYTSPEETITLPTEV 905
QY 389 PAVSGAVNLSLPAVTFEGETSGAICLPYKP-----VSFCMDHICKPVIGTPOIK 441
DB 906 PTVS-TEVTNVSPPEET-----SVPEETITLITTVSPPEETVPPIGTTL-- 949
QY 442 LAQPGVLSOPAGI-----PTGSSSKOLFSLFHVQOPSNGEKOVTTISHSST----- 490
DB 950 ----PTEVLVPIVTFEPTGETT--VPTEVPYVSTENTGHTEVTVVFPETISTIPREAVT 1004
QY 491 -----LTIQKCGOKTMPVN-TTIPTSOFP-----ASILKOITLPGNKILS 530
DB 1005 VLPASIPPEETTPTEVTTTPPETITPAVTVTPPASIPPEETASLLETVTTPPETTTT 1064
QY 531 ----LQASPOKNIKIKENVT-----SCFRDE 552
DB 1065 PTEVTVVPEKTTITPTEVTVTPPASIFPEE 1094

RESULT 10
ID DAN4_YEAST STANDARD: PRT; 1161 AA.
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

```

DE Cell wall protein DAN4 precursor.
GN DAN4 OR YJR151C OR J2223.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_Taxid=4932;
RN (1)
RP SEQUENCE FROM N.A.
RA Scarce T.
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN (2)
RP REGULATION.
RX MEDLINE=21113168; PubMed=11160904;
RA Cohen B.D., Serfil O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
RT manoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites.";
RL Nucleic Acids Res. 29:799-808(2001).
CC -1- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SRP1 / TIR1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
DR EMBL; 249651; CA89684.1; -.
DR SGD; S0003912; DAN4.
DR InterPro: IPR000992; SRP1_TIR1.
DR Pfam: PF00660; SRP1_TIR1.1.
DR PROSITE; PS00724; SRP1_TIR1.1.
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
FT LIPID 1146 1146 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1161 AA; 118358 MW; 7954C15069F0CA58 CRC64;

Query Match 7.1%; Score 195; DB 1; Length 1161;
Best Local Similarity 21.1%; Pred. No. 0.0011;
Matches 149; Conservative 100; Mismatches 242; Indels 216; Gaps 33;

QY 4 VTKVAVSAPKVSQGRLPAPQIVAVKAPNTTITQFPAQLPPTVLIKNSGFL--- 60
DB 369 VTSAAEPTTVSEVTSVEPTRSSQVTSAPPTVSEPTVSVE-PTRSSQVTSAPPTVTS 427
QY 61 -----MLVSPQQTVAETTSNT-----TSRAVPANPQVTKCT---VNSSS 101
DB 428 EFTSSVEPTRSSQVTSAPPTVSEPTVSVEPTRSSQVTSAPPTVSEPTVSVEPTRSS 487
QY 102 QLIKVAVTPVKRLAIGTTVTVTPKPSVQSVAVPTVTPS-KPLNT--VTTLP 158
DB 488 QVTSAAEPTTVSEF-----TSSVEPI-RSSQVTSAPPTVSEPTVSVEPTRSSQVTTTP 542
QY 159 -SSLGAS--STPSNP-----NLKAENSAVOINLS-----PTMLENVKCKKNFLAML 203
DB 543 VSSFSGTFSSETTSAPLPSKATTAESISSNQIITSSBLIYSSVITSSSETPSSIEVL 602
QY 204. IRLAGSGSGPEMGQVKKVLEQLLAKIEAE-----FTRKLVLELSSQPHLYVP 256
DB 603 TSSGISSVEP-----TSLVGPSSDESISSTESLATSTFTSAVVSSKAA-----DF 650
QY 257 LKRSVALROLPLNSQFIQOCVQOQTSMDVAVTATCTTVTTSFVTVTT----- 304
DB 651 FTRSTVSAKSDVSGNS-----TQSTTFEFTPTPLAVSTVTVTSSDVSPIRIPS 702
QY 305 -VSSSGSEKSIIVSGAT--APRTVSQ-----TLNPLAGPVGAKAGVYTLHSVGP 351

```



```

Db 703 EISSPSSSTAIITSTSTFLAERTSSLYLSSNNSSFFLSTFT---VSQISVSSFSMEP 758
QY 352 TA-----ATGTTAGTGLLOF-----367
Db 759 TSSVASFSSSPLVSSRSNCSQARSSTISSGLPSTIENRNATSTFTNLSTDEIVITS 818
QY 368 -----SKPLVTSANVTVTVSLQDEKPVSGTAVALTSLPAVTFG-----406
Db 819 CKSSCTMEDSVLTKTVQSVETFTTSCS-----GICITTLMSPTTINAKANTLITTI 870
QY 407 ETSAAICLPKVRKPVSGFCHMDHCK-----PYIGTPVOIKLAQPGPVLSQAPGIFPGSSSK 462
Db 871 ETS-----TVEFTTTCPCGVCSLTLPVY--FTTISEATTTATISCEDNEEDITSTET 921
QY 463 QLFSLFHVQOOPSGG-----NEKQVTTISHSTLTIOK-----CGQKTM 501
Db 922 ELALLETITSCSGGICITTLMSPTTINAK-ANTLITTEISTVETFTITTCGCGVCSILTV 980
QY 502 PVNTIIPTSQPPASILKQITLPGNKILSLQASPTQKRNRIKENVTSC 548
Db 981 PVNTI--TSE--ATTATISCEDNE-EDVASTKTELLTMTTITSC 1021

```

```

RESULT 11
YAG3_YEAST STANDARD; PRT; 1322 AA.
AC P39712;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 138.1 kDa protein in FLO9-GDH3 intergenic precursor.
GN YAL063C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -1- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U12980; AAC04971.1; -
CC SGD: S0000059; YAL063C.
CC InterPro: IPR001389; Flocculin.
DR Pfam: PF00624; Flocculin; 13.
KW Hypothetical protein; Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 24
FT CHAIN 25 1322 POTENTIAL.
FT TRANSMEM 366 388 HYPOTHEICAL PROTEIN YAL063C.
FT TRANSMEM 754 775 POTENTIAL.
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 644 644 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 689 689 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 734 734 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 888 888 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1322 AA; 138072 MW; ADQFDLFI3267CEA CAC64;

```

Query Match 7.18; Score 195; DB 1; Length 1322;
 Best Local Similarity 20.18; Pred. No. 0.0013;
 Matches 118; Conservative 91; Mismatches 219; Indels 158; Gaps 25;

```

QY 22 LPAPQ-IVAAVAPMTTIOF-----PANQLPQGVILKSNNG-----PLMVSPOQT 68
Db 663 LPDEETIIVIRPTTATATAMTTQPMWDTSTSTETVTVGNGLPDEETIIVIRPTT 722
QY 69 VTRAET-----SNITSRPAPAPQVICKTVPNSSQLIKKAVTPVK 113
Db 723 ATAMTTQPMWDTSTSTETVTVGNGVPD-ETVIVIRPTSE-----768
QY 114 KLAQIGTTVTVPKPSSVQSAVPTSVTVTPCKPLN-TVYTIK-PSLIGASSTPSNP 171
Db 769 -----GLISTTEEPWGTFTSTETMTVGTNGQPDDEVIVIRPTSEGLVTT-TEEP 822
QY 172 -----NKKANSAAVQINLSPTMLENKKRKNFLAMLR--LACSSQSPENQGVKK 222
Db 823 WTGFTSTSTETVTVGNGQPT-----DETVIVIRPTTATISSLSSSSGO---- 869
QY 223 LVEOLLDAKIEAEFTKRLVLELKSPPQPLVPLKRSVALROLPLNSQSFIOQVOQT 282
Db 870 -----ISFTSARPIITPPPS-----NGTIVISSVYS 900
QY 283 S--SDMVIATCTT--VYTSVTVTVVSSQSEKSIIVSGATAPRTVSQTLNPLAPVG 338
Db 901 SDTSLIVISSSVTSVTSVSPVSSIFSSP-----VISTTTSASILSES----- 946
QY 339 AKAGVVLHSGPTPAAGCTAGTGLQTSKPLVTSVA--NTVTVSLQPKPVSGTAV 396
Db 947 SKSSVITPSS--STGSESESTGSASSSSSSISSEPKSTYSSSLPPTVATSTQEI 1004
QY 397 TLSLPAVTFEFGAALCLPSVKPVSGFCHMDHCKPVIGTV-----QIKLAQPGVLSQ 452
Db 1005 TSSLPVYTTKTSBOTIV-----TVTSCSEHCVTESLSAIVSTAVTVSGATTEYTW 1059
QY 453 AGIPGSSSKQLFSLFHVQOPSGNEKQVTTISHSTLTIOK-----GQ 498
Db 1060 CPSTTEITKQ--TTEYTKQTKCTEQTETTKQTLVTVTSSCEDVCSKTASPAIVST 1116
QY 499 KTMVNTI-----IPTSQPPASILKQITLPGNKILSLQASP 535
Db 1117 STAVINGTTEYTTWCPISTTESKQQTLLVTVTSCGSGVSEETSP 1162

```

```

RESULT 12
ZAN_HUMAN STANDARD; PRT; 2700 AA.
AC Q9Y493; 000218;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin (Fragment).
GN ZAN.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-2379 FROM N.A.
RX MEDLINE=99018118; PubMed=9799793;
RA Glockner G., Scherer S., Schatevov R., Boright A., Weber J.,

```

RA Tsui L.C., Rosenthal A.:
 RT "Large-scale sequencing of two regions in human chromosome 7q22:
 RT analysis of 650 kb of genomic sequence around the EPO and CUGL1 loci
 RT reveals 17 genes.";
 RL Genome Res. 8:1060-1073(1998).
 RN [2]
 RP SEQUENCE OF 2338-2700 FROM N.A.
 RC TISSUE-Testis;
 RA MEDLINE-97271566; PubMed-9126492;
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN).";
 RL Genomics 41:119-122(1997).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE Egg. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVUDUCTAL ISTHMUS.
 CC -1- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF053356; AAC78790.1; -;
 DR EMBL: U83191; AAC51208.1; -;
 DR Genev; HGNC:12857; ZAN.
 DR MIM: 602372; -;
 DR InterPro: IPR0000561; EGF-like.
 DR InterPro: IPR000998; MAM_domain.
 DR InterPro: IPR002919; TIL_Cysrich.
 DR InterPro: IPR003328; TILa_Cysrich.
 DR InterPro: IPR001846; VWF_D.
 DR Pfam: PF000094; vwd; 4.
 DR Pfam: PF00629; MAM; 4.
 DR Pfam: PF01826; TIL; 5.
 DR Pfam: PF02345; TILa; 4.
 DR SMART: SM00216; VWD; 1.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS50060; MAM_2; 4.
 KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
 FT NON_TER 1
 FT DOMAIN <1 109 MAM 1.
 FT DOMAIN 112 136 MAM 2 (PARTIAL).
 FT DOMAIN 161 326 MAM 3.
 FT DOMAIN 322 446 MAM 4.
 FT DOMAIN 483 951 66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
 FT DOMAIN (MUCIN-LIKE DOMAIN).
 FT DOMAIN 953 1065 VWFD 1 (PARTIAL).
 FT DOMAIN 1066 1454 VWFD 2.
 FT DOMAIN 1455 1861 VWFD 3.
 FT DOMAIN 1862 2292 VWFD 4.
 FT DOMAIN 2293 2684 VWFD 5.
 FT DOMAIN EGF-LIKE.
 FT CARBOHYD 74 74 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 403 403 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1023 1023 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1099 1099 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 1618 1618 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1737 1737 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1832 1832 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1878 1878 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2136 2136 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2505 2505 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 2374 2379 N-LINKED (GLCNAC...) (POTENTIAL).
 FT NON_TER 2700 2700 NNOKMA -> RAGPGP (IN REF. 1).
 SQ SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;
 Query Match 7.1%; Score 194.5; Db 1; Length 2700;
 Best local similarity 22.7%; Pred. No. 0.0033;
 Matches 138; Conservative 84; Mismatches 235; Indels 151; Gaps 33;
 QY 1 GLTVTRV-APVAPP-----KVSQPRLPAPQIV-AVRAPPTTTI 38
 DB 372 GTMLELLGSPASPPIPLMKRVGSQRPYQNTSVTPSGHQPMQLIFKIGSGNTASV 431
 QY 39 QPANIQLPPGIVLINSQPLMLVSPQCTVTAETTSNITSRPVAVPANDVYKICTVPR 98
 DB 432 VANGFLLINGTCPRVY-LPELPVSPVSTGSETT-GLTENPTISTKRPVTS----- 483
 QY 99 SSSQLIKVAVTPVKKLAQIGTVTTVPKSSVQSAVAVTSVTVYPGPLNTVTLTP 158
 DB 484 -----TEKSVTEK-----PTVRK-----EKPTIPEKPTISTEK- -TIPSEK 522
 QY 159 SSLGASTSPNEPLKNAENSAVOINLSPTMLENVKCKNFLMLIKLCSGSPSEMGQ 218
 DB 523 -----NMPSEKPTISE-----KPTIPEKPTIS-----EKPTIPEKPTI- 559
 QY 219 NVKKIVBQLDAKIEAEFTKRIYELKSSPOHLVPLFKSVVALROLPNQSTI--Q 276
 DB 560 STEKPTVPEEPTPEETT--TYME-----EP-VIPEKPSIPTKPSIPTKPSIME 611
 QY 277 QCVQOTSDDVIATCTTVTTSPTVTTSSQSEKSIISG-----ATAPRYS-----V 327
 DB 612 ETTISTEKPICEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI 671
 QY 328 QTLNPLAGP--VGAKAGVTLHSVGPATAGTGTAGTGLQTSKPLVTSVANTVT- -VS 383
 DB 672 PTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI-----ISTEPTPEETIISTEKS 726
 QY 384 LQPEKPVSTANTVLSBANTPEBTSGAACLSVAPVVSFCMDHICKVIG- ------P 437
 DB 727 IPEKPTLPTEETTSVEETI--STEKLTIPMEKPTIS-----TEKPTIPEKPTISP 778
 QY 438 VQI-----KLAQPG-----PV-----LSQPAQIPGSSSKOLFSLFHVQOPSQGN 478
 DB 779 EKLTIPTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI 834
 QY 479 EK-----QVTTISHSSTLTIQKCGQRTMPVNTIIPTSQPPASILKQTLPGNKILSLQA 533
 DB 835 EKPITPEETIIS-TEKLTIPTEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI 887
 QY 534 SPTQKNRI 541
 DB 888 IPEKPTI 895
 RESULT 13
 VGS0_HSV11 STANDARD; PRT; 670 AA.
 AC 000130;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical gene 50 protein.
 GN 50.
 OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Ictalurid Herpes-like viruses.
 NCBI_TaxID=10401;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-Auburn 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davidson A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M75136; AAA88153.1; -
DR PIR; F36791; F36791.
KW Hypothetical protein; Repeat.
FT REPEAT 143 158
FT REPEAT 171 186
FT REPEAT 200 214
FT REPEAT 215 233
FT REPEAT 234 252
FT REPEAT 253 268
FT REPEAT 279 293
FT REPEAT 294 309
FT REPEAT 320 334
FT REPEAT 335 349
FT REPEAT 362 376
FT REPEAT 377 391
FT REPEAT 392 406
FT REPEAT 407 421
FT REPEAT 422 436
FT REPEAT 437 452
FT REPEAT 464 477
FT REPEAT 478 493
FT REPEAT 504 517
FT REPEAT 518 531
FT REPEAT 532 545
FT REPEAT 546 559
FT REPEAT 560 573
FT REPEAT 574 587
FT REPEAT 588 601
FT REPEAT 602 615
FT REPEAT 616 629
SQ SEQUENCE 670 AA; 64174 MW; 2B64A781C519EB84 CRC64;
Query Match 7.0%; Score 193.5; DB 1; Length 670;
Best Local Similarity 22.1%; Pred. No. 0.00069;
Matches 136; Conservative 60; Mismatches 276; Indels 143; Gaps 27;
QY 1 GTLVTKAAPSAPPK-----VSSGPRLPAPQI--VAVKAPNT----- 35
DB 62 GDAVTGQOPVERESSLPRTTPSLAHITISKMISLGR-PPPTIPGCVTTTPNDAPVP 120
QY 36 -----TTTQFPANLQLPQGTVLKSNSSGRLMVS-----OOTVTRA 72
DB 121 GSVHTTARVVDTITTKOTPTTPATPAGA---NDTANITTTATPAGANDTANITATPAGA 176
QY 73 ETTSNITSRAVPANPQ--TVKICTVNSSQLKKAATPVVKLAQIGTVVTVTP-KP 129
DB 177 NOTANITTTATPAGANDTAVTTTSATPAGANDTAVTTTPATPAGANDTANITVTTTPAMP 236
QY 130 SSVQSVAVPVSVVTPGCKPLN-----TVTLKPSLSGASSTPANEPLKKAENSAVAQIN 184
DB 237 AGANDTANGTA-VTTTPAMPAGANDTANITATPT--GANDTANVTMAGATDIV---VT 290
QY 185 LSPMLENVKCKNPLAMLILACSSGSPMGQNVKKLVQDLDAKIEAEFTKLYVE 244
DB 291 TTPAMPAGANDTAN-ITTATPAGANDTANVTMPAGATDVVTTTPAMPAGANDTANVTKP 349
QY 245 LKS-----SPQHLVPLPKSVVALRQLLPNSQSFIOCVQGTSSDVIATCTTVTTS 299

```

```

DB 350 AGSTDTVTTTPAMPAGATDVTTTPAMPATGAT--DTVVTTTPAMPAGATDVTTTP 406
QY 300 -----VYTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGVCAKAGVYTLHSVG 350
DB 407 AKPAGANDTAVTTTPAMPAGANDVTVTTATPATPAGANDTAN-VTKPTGATDVTTTATVK 465
QY 351 PPAATGCTTAGTG-----LLQTSKPLVTSVANTVTVSSLP-----EKPV-V 391
DB 466 PTGATGVTTTAKPTGANDTANVTKP--TGATGVVTTTAKPTGANGTVVATAKPTGA 523
QY 392 SGTAVTSLPVAVTGEGTGA--AICLSVKRPVSPCMDHICKPVIQIPVQIKIAQCPVL 449
DB 524 TGVVTT-----TAKPTGANGTVTTTAKPT-----GATGVTV 556
QY 450 SOPAGIPTGSSSKQLESLFHVQOPSGGENEKQVTTISH--SSTLPIQCGQKTMVNTII 507
DB 557 TTTAK-PTGANG-----YTTTAKPAGANGVTTTAKPAGANGTVTTTAKPAGANGTV 611
QY 508 PTSQPPASILKQIT 522
DB 612 TTTAKPAGANGTVT 626

```

```

RESULT 14
VGLX_HSVB
ID VGLX_HSVB STANDARD; PRT; 797 AA.
AC P28968;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=9229556; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davidson A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M86664; AB02506.1; -
DR PIR; H36802; VGBEX1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 797
FT DOMAIN 23 465
FT TRANSMEM 766 790
FT CARBOHYD 590 590
SQ SEQUENCE 797 AA; 80342 MW; 50C9ED9211F5E5B2 CRC64;
Query Match 6.9%; Score 189.5; DB 1; Length 797;
Best Local Similarity 18.9%; Pred. No. 0.0014;
Matches 118; Conservative 65; Mismatches 269; Indels 173; Gaps 15;
QY 1 GTLVTKAAPSAPPKVSSGPRLPAPQIVAAKAPNTTIIQFPANLQLPQGTVLKSNSSGPL 60
DB 45 GTTSSSPPTSPPTSSP-----PTSTHTSPSSTSTQ----- 79
QY 61 MLVSPQGVPAETTSNITSRAVPANPQVTKICTVNSSQLKKAATPVVKLAQIGT 120
DB 80 ---SSSTANISSAPSTASSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 133

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 38.77 Seconds
(without alignments)
2933.659 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_552

Perfect score: 2758
Sequence: 1 GILVTKVADVPASAPKVSQGP.....ASPTQKRIKENVTSQFIDE 552

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTRMBL_21: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archaeal: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	558	20.2	662	11	O91W66 mus musculus
2	234.5	8.5	851	5	O8T9E0 drosophila
3	233.5	8.5	528	6	O29071 sus scrofa
4	226	8.2	1275	5	O76602 caenorhabditis
5	224.5	8.1	1349	4	O8W044 homo sapien
6	222.5	8.1	1079	5	O9N4S7 caenorhabditis
7	222	8.0	1979	11	O9QY35 mus musculus
8	219.5	8.0	1332	4	O8TE50 drosophila
9	217	7.9	2045	11	O61191 mus musculus
10	215	7.8	2045	11	O9QW42 mus musculus
11	214.5	7.8	1795	5	O76602 drosophila
12	214	7.8	1029	4	O8T9E0 drosophila
13	212.5	7.7	2232	5	O91365 caenorhabditis
14	212	7.7	842	5	O9VGC9 drosophila
15	212	7.7	864	5	O95045 drosophila
16	212	7.7	864	5	O9VGC8 drosophila

17	208.5	7.6	800	3	O8TFG4 schistosoma
18	208	7.5	1844	5	O22579 caenorhabditis
19	206.5	7.5	1203	5	O9N5K0 caenorhabditis
20	205.5	7.5	1246	4	O15052 homo sapien
21	205.5	7.5	2382	4	O9H4A3 homo sapien
22	204.5	7.4	645	4	O9H048 homo sapien
23	204.5	7.4	1324	4	O15312 homo sapien
24	204	7.4	2187	11	P70670 mus musculus
25	201.5	7.3	961	3	O92223 emulicella
26	201.5	7.3	5374	11	O99NDO mus musculus
27	201	7.3	873	10	O9C548 arabidopsis
28	199.5	7.2	513	4	O43418 homo sapien
29	199.5	7.2	2601	4	O9B284 homo sapien
30	199.5	7.2	2601	4	O9B186 homo sapien
31	199.5	7.2	2624	4	O9B285 homo sapien
32	199.5	7.2	2624	4	O9B287 homo sapien
33	199.5	7.2	2689	4	O9B287 homo sapien
34	199.5	7.2	2689	4	O9B185 homo sapien
35	199.5	7.2	2721	4	O9B283 homo sapien
36	199.5	7.2	2721	4	O9B190 homo sapien
37	199.5	7.2	2724	4	O9B288 homo sapien
38	199.5	7.2	2724	4	O9B288 homo sapien
39	199.5	7.2	2812	4	O9B286 homo sapien
40	199.5	7.2	2812	4	O9B188 homo sapien
41	197	7.1	648	4	O14760 homo sapien
42	197	7.1	2971	4	O9Y519 homo sapien
43	196.5	7.1	867	12	O39782 equine herp
44	196	7.1	3507	5	O23587 caenorhabditis
45	195	7.1	826	12	O8V015 equine herp

ALIGNMENTS

RESULT 1

AC O91W66 PRELIMINARY; PRT; 662 AA.
ID O91W66
AC O91W66; 01-DEC-2001 (TREMUREL_19, Created)
DT 01-DEC-2001 (TREMUREL_19, Last sequence update)
DT 01-JUN-2002 (TREMUREL_21, Last annotation update)
DE TANA-binding protein associated factor TAF1135 (Fragment).
GN TAF4A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=PO BRAIN;
RA Metcalf M., Brunkhorst A., Neuman T.;
RT "Cell Type Specific Expression of the TFIID Component TAF1135 in the Nervous System";
RL Exp. Cell Res. 0:0-0(2001).
DR EMBL; AY038601; AAK94779.1; -
DR MGD; MGI:2152346; Tafa4.
FT NON_TER
FT SEQUENCE 662 AA; 71398 MW; 91A75F38CBDD0DA4 CRC64;

Query Match 20.2%; Score 558; DB 11; Length 662;
Best Local Similarity 31.1%; Pred. No. 2.7e-26;
Matches 171; Conservative 67; Mismatches 142; Indels 170; Gaps 20;

OY 20 PRLAPQIVAVKANTTIOFPAKOLPFGIVLTKNSGPKMLVSPQOTV-----TRA 73
DB 10 PRLPQP-----QKPN-NIO---NFOLPPGVAVLRSSNGQLML-PQDALAQMOMAHQAQ 59
OY 74 TTSNITSPAPVAPNQTIVCTVNSSSQLTKKVAVPVKLAQIGTIVTTVPKPSVQ 133
DB 60 PQSMARPRAPPTAPVQISTVQAPGPTIAR-QVTP-----TTIIVQV---SQAQ 107
OY 134 SVAVPSTVVTVPKPLNTVT--TLKPSISGASS-----TPSNE-PMUKENSAAVOINTL 185

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platzer M., Brenner V., Reichwald K., Wiene T., Oksche A.,
RA Rosenthal A.;
RT "Comparative sequence analysis of the mouse L1cam locus and the
RT corresponding region of human Xq28";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF133093; AAF22156.1; ;
DR MGD: MGI:105942; Hcfc1.
DR MGD: MGI:1349215; Abcd1.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 4.
DR SMART: SM00060; FN3; 1.
FT NON_TER 1
SQ SEQUENCE 1979 AA; 203436 MW; ABE98C5DDE688A9 CRC64;

Query Match 8.0%; Score 222; DB 11; Length 1979;
Best Local Similarity 23.2%; Pred. No. 3.5e-05;
Matches 140; Conservative 71; Mismatches 232; Indels 160; Gaps 29;

QY 4 VTAAVAPVAPKVSQPRPAPQIYAVKAPNTTIOFPANL-----QLPQGVV-- 52
DB 417 VPAVLKVTGPQATGTGPLVTRKRPASQ--AGKAPYVTV--SLPASYAMVVPVTSOAGTIVS 473
QY 53 -----IKNSGRLMLVSPQOTVTRAEFTSNITSBPANPANOYTKI 93
DB 474 NPQMSGAALAAATAATOKIPSSAPVLSVPAGT-----IVKTVAVTGGTTTLPATYKV 529
QY 94 CTVP-----NSSOLIKKVA--TPVKLAIQIGTVVTVTPKSSQVAVPTSVVTVTP 146
DB 530 ASSVAVSNPATRLKLTAAQVGTSSAANTSTRPITTVKSGTV--IVAAQAOVTVTV 588
QY 147 GKPLNTVTVTLK--PSSL--GASSTPSNEPNL-----KAENSAVOINLSPTMLENKKCN 198
DB 589 GGVTKTTTLVKSPIVSPGSGALISNLGKVMGVOTKPVQTSAVTGOASTGCVTOIIQTKG 648
QY 199 FL--AMLIKILACSGSQSP-----EMGQNVKKLVBQLDAKIEAEFTKRLVYELKSS 248
DB 649 PLRPGTILKLTVSADGKPTTITTTTQASGAGTKPTILGI-----SS 689
QY 249 PQLPLVPLFKKSVALLRQLLPNSQSFIOQVOQTSMDVIACTTVTVTT-----S 298
DB 690 VSPSTT--KPGTTTILKTIIPMSALITQAGATVTSPIGKSPITITTKWTSGTGAPA 746
QY 299 PVVT-----TVSSSQSKSIIVSGA-----TAPRT-----VSQTLNPLAGPV 337
DB 747 KIITAVPKRIAGHGQGVTVVLKAGAPQPTILRTVPMGCVRLVTPVTSAVKPAVTVTL 806
QY 338 GAK--AGVTLHSVGPPTAATGTTAGTGLQTSKPLVTSVA--NTVTVSLQPEKPVVSG 393
DB 807 VKKTTGVTTLIGTVTGYST--SLAGAGHSTSLATPITTLGTLATLSQ-----VNP 860
QY 394 TAVTLSPAVTFGETSGAICLPKVPKPVSPCMHDICKPVGITPVQIKL--AQPGLVLSQP 452
DB 861 TAITVSAAGTTLTAAGGLTPTITMOPV-----SQPLQVULITAPSGVEAQP 907
QY 453 A-GIPTSSSSKQLSLFHVVOQPSGNGEKQVTTTISHSTLTIOKCGQKTMVNTTIIPSSQ 511
DB 908 VHDLPVS-----ILASP-----TTROPATVITADSGQDVQPGFVTLVCS 948
QY 512 FPP 514
DB 949 NPP 951

DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE EMSY protein.
CN C1IORE30.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hughes-Davies L.;
RT "EMSY is amplified in breast cancer and displays a BRCA2 dependent DNA
RT damage response";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ430203; CAD22881.1; ;
SQ SEQUENCE 1322 AA; 141467 MW; 7F8C95E8BA0FC9F0 CRC64;

Query Match 8.0%; Score 219.5; DB 4; Length 1322;
Best Local Similarity 22.7%; Pred. No. 3.1e-05;
Matches 144; Conservative 88; Mismatches 230; Indels 171; Gaps 30;

QY 5 TKVAPVAPKVSQPRPAPQIYAVKAPNTTIOFPANL--LQPLPG--TVLIKNSGRLM 61
DB 149 TTSPTSTP-----VPSGSIAVTKSPRPAS--PASNMYVLPSSGVYVYKVSQSCDE 197
QY 62 LVSPQOTVTRAEFTSNITSBPANPQTV-----KICVTPNSS-----SOLIKKVA----- 108
DB 198 DEKPRK--RRRTSSSSSPVYKLEVPKAVPVSTTTVVPVSGSKMSINOSIANSILPP 254
QY 109 -VTPVK-----KLAQIGT-----VTVTPKSSQVQV-----AVPTSVTV 144
DB 255 HNSPVKIFKTPSTQGTNTTQKVIIVTSPSTFVNPILSKHMYAVTKVPTSVYAS 314
QY 145 TPCKLNTVTVTLKPSL-----GASSTPSNEPNL-----LKAENSA 180
DB 315 TTQKP--PVYITASQSLVSNSSGSSSTSPSPITPTVAVTAVVSTPSVMSVTAQGVST 373
QY 181 VOINLSPTMLENVKCKNFKMLIKLACGSSQSPMGQNVKKLVBQLDAKIEA--EFTTR 239
DB 374 SAIKMASTRLSPKSLVAPVQIL-----AOFKQHQQSRK--QOLYVOQOQTQOQVAAQ 425
QY 240 KLYVELKSSPQHLVPLFKKSVALLRQLLPNSQSFIOQVOQTSMDVIACTT--TV 295
DB 426 PSPVSHQOQPOQSPLPKIPKPTIQIKQ-----ESCVKIITQOVQSKILPKVTVTLPTS 480
QY 296 TTSPTVTTVSSSQSKSIIVSGAIPRTVSVQTLNPLAGPVGAAGVTV----- 345
DB 481 SNSPTMVVSSMGALMTTKLVTPGTQATYTRPVPSPISGRMAATPGAAIYVKTSGSII 540
QY 346 -----LHSVG-----PRAATGTTAGTGLQTSKPLVTSVANTV--TVSLQPEKPV 391
DB 541 TVPKSLATLIGKTISSNIVSGTTTKITTPKTSKPNVIVOKTKGKTTIOGLPKANV 600
QY 392 S-----GTAVTLSPAVTFGETSGAICLPKVPKPVSPCMHDICKPV--IGTPVQIK-- 441
DB 601 TTLNAGGEKTIQVVP-----TGAKPAIILATRPITKMI--VTPQGISGTVQPAKI 651
QY 442 -----LAQGPVLSQPAIGIPGSSSKQLSLFHVVOQS----- 475
DB 652 IPTKIVGQOQKTVLILKPKVTFQATVV--SEQTRQVTV--ETLQASRVAEAGNSSIQ 707
QY 476 -GNGEKQVTTISHSTLTIOKCGQKTMVNTII 507
DB 708 EGKEPQNTDSSSSSTSSSQSDSQFVYVHI 740

RESULT 8
Q8TE50
ID Q8TE50 PRELIMINARY; PRT: 1322 AA.
AC Q8TE50;

RESULT 9
O61191
ID O61191 PRELIMINARY; PRT: 2045 AA.
AC O61191;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Transcription factor C1 (HCF).
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Kristie T.M., Dasher R.;
RT *cdnas encoding the mouse homolog of the human transcription factor C1
RT (HCF).
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U03925; AAB0163.1; -.
DR MGD; MGI:105942; Hcf1.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 5.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210537 MW; B89CC2FDA35B969F CRC64;

Query Match 7.9%; Score 217; DB 11; Length 2045;
Best Local Similarity 23.5%; Pred. No. 7.4e-05;
Matches 141; Conservative 76; Mismatches 224; Indels 158; Gaps 31;
QY 7 VAPVSAPPKVSQPRRL--PAQIVAVKAPNTTIOIPANLQIPGGIVLILKSNSGPLMLV 63
DB 486 VLKVTGQATGTPPLVIMRPAQ--AGKAPVTV-SLPASVVR--VVPQSAQGTIVIGS 539
QY 64 SPQQT-----VTRAETTSNT--TSRP--AVPAN-----PQIVKICTVP 97
DB 540 NPGMSGMAALAAAAAATQKIPSSAPTMVSPAGTTIVKTVAVTPGTTLPATVAVKVASP 599
QY 98 -----NSSOLIKKAV--TPVKKLAQIGTVTVTPKPSVQSAVAPTSVTVTPGKPL 150
DB 600 VMSNPNATRLKTAAGVGSVSSAANTSTRPIITVHKSGTV--TVAQAQVTVTVVGGVT 658
QY 151 NTVTTLK-PSL-GASSTPSNEPNL-----KAENSAVQINLSPTMLENKKCKNFL-- 200
DB 659 KITLVKSPISVPGSALISNLGKVMVSVQTKPVQTSVAVGQASTGTVQIIRTKGPLA 718
QY 201 AMLIKLACSSQSP-----EMQNVKKLVQGLDAKLEAEETFRKKLYVELKSSQPH 252
DB 719 GTILKLVTSADGKPTIITTTQASGAGTKPTIIGI-----SSVSPS 759
QY 253 LVPLFKSVVALRQLPNSQSFIOQCVQSSDMVATCTTVT-----SPVVT 302
DB 760 TT---KPGTTTITKIPMSALITQAGATGVTSBGIKSPIITITTKVMTSGTGAPAKIIT 816
QY 303 -----TVSSSQSEKSIIVSGA-----TAPRT-----VSQTLNPLAGPVGAK- 340
DB 817 AVPKIATGHQGOQGVTVVLKGAPGPGTILRTVPMGVRLLVTPVSAVAPATVTLVVGK 876
QY 341 -AGVTVLHSGPTAAGTGTAGTGLQTSKPLVTSVA--NTVTVVSLQPKPVVSGTAVT 397
DB 877 TGGVTLTGTVTGIVST--SLAGAGHSTASLAPITTLTGITATLSQ-----VINPFAIT 930
QY 398 LSLPAVTFGTSAGALCLPSVKPVVSCMDHICKPVIGTVQIKL-AQPGVLSQPA-GI 455
DB 931 VSAAGTTLTAAGGLTPTITMOPV-----SPTQVTLITAPSGVEAQPVHDL 977
QY 456 PTGSSSKOLFSLFHVVOQPSGNEKQVTTISHSSTLTIOCKGQKTMPTVNTIIFTSQPP 514
DB 978 PVS-----ILASP-----TTQPTATVTVIADSGGDVQPGTVTLVCSNPP 1017

01-MAY-2000 (TReMBLrel. 13, last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE C1 transcription factor.
GN HCF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Kristie T.M.;
RT *cdnas encoding the mouse homolog of the human transcription factor C1
RT (HCF).
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U08082; AAD09225.1; -.
DR MGD; MGI:105942; Hcf1.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 5.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210546 MW; 7AD3BFCDD78EABA9C CRC64;

Query Match 7.8%; Score 215; DB 11; Length 2045;
Best Local Similarity 23.5%; Pred. No. 9.8e-05;
Matches 141; Conservative 75; Mismatches 225; Indels 158; Gaps 31;
QY 7 VAPVSAPPKVSQPRRL--PAQIVAVKAPNTTIOIPANLQIPGGIVLILKSNSGPLMLV 63
DB 486 VLKVTGQATGTPPLVIMRPAQ--AGKAPVTV-SLPASVVR--VVPQSAQGTIVIGS 539
QY 64 SPQQT-----VTRAETTSNT--TSRP--AVPAN-----PQIVKICTVP 97
DB 540 NPGMSGMAALAAAAAATQKIPSSAPTMVSPAGTTIVKTVAVTPGTTLPATVAVKVASP 599
QY 98 -----NSSOLIKKAV--TPVKKLAQIGTVTVTPKPSVQSAVAPTSVTVTPGKPL 150
DB 600 VMSNPNATRLKTAAGVGSVSSAANTSTRPIITVHKSGTV--TVAQAQVTVTVVGGVT 658
QY 151 NTVTTLK-PSL-GASSTPSNEPNL-----KAENSAVQINLSPTMLENKKCKNFL-- 200
DB 659 KITLVKSPISVPGSALISNLGKVMVSVQTKPVQTSVAVGQASTGTVQIIRTKGPLA 718
QY 201 AMLIKLACSSQSP-----EMQNVKKLVQGLDAKLEAEETFRKKLYVELKSSQPH 252
DB 719 GTILKLVTSADGKPTIITTTQASGAGTKPTIIGI-----SSVSPS 759
QY 253 LVPLFKSVVALRQLPNSQSFIOQCVQSSDMVATCTTVT-----SPVVT 302
DB 760 TT---KPGTTTITKIPMSALITQAGATGVTSBGIKSPIITITTKVMTSGTGAPAKIIT 816
QY 303 -----TVSSSQSEKSIIVSGA-----TAPRT-----VSQTLNPLAGPVGAK- 340
DB 817 AVPKIATGHQGOQGVTVVLKGAPGPGTILRTVPMGVRLLVTPVSAVAPATVTLVVGK 876
QY 341 -AGVTVLHSGPTAAGTGTAGTGLQTSKPLVTSVA--NTVTVVSLQPKPVVSGTAVT 397
DB 877 TGGVTLTGTVTGIVST--SLAGAGHSTASLAPITTLTGITATLSQ-----VINPFAIT 930
QY 398 LSLPAVTFGTSAGALCLPSVKPVVSCMDHICKPVIGTVQIKL-AQPGVLSQPA-GI 455
DB 931 VSAAGTTLTAAGGLTPTITMOPV-----SPTQVTLITAPSGVEAQPVHDL 977
QY 456 PTGSSSKOLFSLFHVVOQPSGNEKQVTTISHSSTLTIOCKGQKTMPTVNTIIFTSQPP 514
DB 978 PVS-----ILASP-----TTQPTATVTVIADSGGDVQPGTVTLVCSNPP 1017

RESULT 10
Q9QWH2 PRELIMINARY; PRT; 2045 AA.
AC Q9QWH2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)

RESULT 11
Q76894 PRELIMINARY; PRT; 1795 AA.
AC Q76894;
DT 01-NOV-1998 (TReMBLrel. 08, Created)

```

; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aljun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match 7.9%; Score 219; DB 10; Length 5179;
Best Local Similarity 21.9%; Pred. No. 5.8e-06;
Matches 136; Conservative 63; Mismatches 227; Indels 194; Gaps 29;

OY 2 TLVTAVAPASAP-----PKVSSGRLPAPQIVAKAPMTT-----QEPAN 43
Db 3765 TTTTIVTPPTPTGQTPTPTTPTTTTIVTPPTPTGQTPTTTTIVTPPTPTG 3824
OY 44 LQLPPTVLKNSNSGPLMLVSPQGTVAETTSNTS-----RPAVPAPOTKICT 95
Db 3825 TQTPPTPTTTTTT---VTPPTPTGQTPTPTTTTIVTPPTPTGQTPTPTTPT 3881
OY 96 ---VPNSSQILKVAATPVKKLAQIGTIVTVPKSSVQS -AAVP-TSVVYTP--- 146
Db 3882 TTTTTPPTPTGQTPTPTTPTT-----TTTTPPTPTGQTPTPTTPTTPTPT 3936
OY 147 -----GKPLNTVTTTKPS-SLGASSTSPNEPLKAKNSAAVOINSPMLKNNCK 197
Db 3937 PTGQTPTPTTPTTTTIVTPPTPTGQTPTT-----ITTTTIVTPPT----- 3983
OY 198 NFLAMLKILACSGSSPEMGQNVKRLVEQLDAKIEAEFTKRLLYELKSSFOPLVPL 257
Db 3984 -----TGTQTP-----TTPITTTTIVTPPT----- 4004
OY 258 KKSVALHQLPLNSQSFIQCVQOFTSDMVATCT-----TSTSPTVTTTYS 306
Db 4005 -----TPTG-----TQTPPTPTTTTIVTPPTPTGQTPTTPTTIVT 4046
OY 307 SSQSEKSIIVSGATAPRTVSVOTLNPLAGPVGAK---AGVTLHVSQPTAATGCT--- 358
Db 4047 TPTPTPTGQTPTTPTTPTT-TVAPPTPTGQTPTTPTTPTTTTIVTPPTPTGQTPT 4105
OY 359 ---TAGIGLQTSPL-----VTSVANTVTVLSQP-----EKPVSGTAVTISL 400
Db 4106 TPTTTTIVTPPTPTGQTPTTPTTPTTTTIVTPPTPTGQTPTTPTTPTTIVT-PT 4164
OY 401 PAVTGETSGAICLPSVKPVVSCFMDHCKPVIQTPQIKLACGPVLSQAPAGIPTSS 460
Db 4165 PPTPTGT-----PTTTPTT-----TTVTP-----PTPTGTQ-TGPTTHTS 4202
OY 461 SKQFLSLHVVQOPSGENKQVTTTSHSSTLTQCGKTMKPVNTIIP---TSQFPAS 516
Db 4203 TAPIEL--TTSNPPESTPQTSRSTSSPLT-----ESTLISLPAIEWTSTAPST 4255
OY 517 ILKQITLPGNKILSLQASPT 536
Db 4256 PTAPTTSGHTLSPPSTT 4275

RESULT 5
US-09-833-263-1068
; Sequence 1068, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aljun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
```



```

QY      9  PVSAPPKSSGGRLEPAPQIVAAKAAENTTICPPANIQLPFGVLIKSNGPLMVS--PQ 66
Db      1841  PVTPE-----SDPSIIFILPLPSVA-----AKLSPVAGGGLPHQSPPKVEMILTR 1887
QY      67  QVTTRAETTSNITSRPAPVAPNQPVIKCTVPNSSQLIKKVAVP--YKLAQIGTTVT 124
Db      1888  QEEPAQSTPS-----PALPPDTRKASDVDDSSSTL--RKILMDKKYVATSVTSVYT 1939
QY      125  TVPKRSSQSVAVAPISVTVTPGKRLNTVYTLIKPSLSGASFPNSNPNLKAENSAVOIN 184
Db      1940  AIAEVSAPAPCLHEAPPPPVDSKKPLEEKTA--PVTNNSEIQASEVLAAAKREKAPV- 1996
QY      185  LSPTMLENVKKCKNFIAMLIKILACSGSOSPENGQVWKKLVEQLIDAKIEAEFEETRKLIVE 244
Db      1997  IAPKITSVISR-----MPVSIIDLENSQKITLAKAPQILGLVSA----- 2036
QY      245  LKSSQPHLVLPPLKKSVALROLLENFSOSFIOCCVQOQSSDMVIACTCTVTTSP--VVT 302
Db      2037  LTGLVNVSLVP-----VNALKGPVKGSVYTLTSLVSTPAGPVNVILKGPVNVTLGPVNVTL 2091
QY      303  TTVSSSQSEKSIIVSGATAPRVSQVQIINPLAGPVGAKGVYTLHSVGTANNGTACT 362
Db      2092  TPVNT-----VGVNAAPGTV-----NAASAVNATASATVTRAGAVTAASGGTATY 2140
QY      363  GLLQTSKPLVTSVANTVTVSSLOPEKPVVSGTAVTSLPAVTFEGSGAICLPSPKPV 422
Db      2141  GTV-----TMAQAVIAPSTR----- 2155
QY      423  SFECMDHICKPYIIGTVQIKLAPG--PVL-SOPAGIPTGSSSKOLFSEFHVYQOPSGNE 479
Db      2156  -----KORASANENSRR--HFGMPVYIDRPADAGSGAGLR-----VNTSEG-- 2196
QY      480  KOVTTIHSSTLTIGCGOKTM-PVNTIIPTSQFPASILKQITLPGNKILSLQASPTOK 538
Db      2197  --VVLIS-----GQKTEGPORISAKISQIPPASAMD-----IEFOOS-VSK 2236
QY      539  NRIK-ENYTS 547
Db      2237  SOVKPDSVTA 2246

RESULT 9
US-09-862-027-40
; Sequence 40, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
;
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: No. US20020142428A1el Kinases and Uses Thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862, 027
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 09/345, 473
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 1601
; TYPE: PRT
; ORGANISM: C. elegans
; US-09-862-027-40

```

```

Query Match      6.4%: Score 176.5; DB 10; Length 1601;
Best Local Similarity 20.6%: Pred. No. 0.00082;
Matches 123; Conservative 93; Mismatches 244; Indels 137; Gaps 28;

QY      5  TKVAVSAPRPKSSGRLPAPQIVWKA1RTTIOFANLQJPGTVLKSNSGRLMYS 64
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      514  TP1PT1PA1TPH-SSAQQQ1IP1PL1TQ-----TSAETIQSSAQGP-----SVPTYMTA 559

QY      65  --PQQTVA1RAETTSNITS1PAVPANPQ1U1CTV1ENSSQSL1LKKVAV1PVKK1LA1IGTV 122
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      560  NIPAS1SP1SAQGP1ULS--PTSA1AV1VP1TM1LHV1K1PS1IPQ1NAV1TT-----AA1PA1AN1N 613

```

```

QY 123 VTYVPKESVQSAVAVPT-----SVYTVPGKPLNVTTL-----KPSLIGASS----165
Db 614 VPPSPAPKPTEDIDTTPILAQNTVYERTISTDASGLVINTPASIASPRAPSATDVASTTAP 67373
QY 166 -TPSNEPNLKAENSAVAOINLSPTMLENVKCKKNFLMLIKLACSGSQ----SPENGON 219
Db 674 VTPAPPTPTTTTDDGAAAAASTTTEKKEKRKSKNKRYMEI-LGDESRNFALVSCRIDTS 73232
QY 220 VKKIVLEOL-----LDATKEEFETPRKLYVLEKSSPPQHVPPL-----KKS-VALR 2655
Db 733 HKSVTFQAPETDKPCTTATATKLLEEDCLKNHVAIVEAOLGEYTIQILNSDGKKGVGTKLA 79222
QY 266 QLL-PPNSOS-FIOOCVOQTSSDAVIAITCTTYYTSPVTT--TVSSSQEKSIIYSGAT 3320
Db 793 TVLDPNSTEPPRTIIVAMPKDS---AADASNTKRIEIEKTPPTPRASQSPNNVQV---T 8466
QY 321 APRVYVQTLNPLPAGPVGAKAGVYTLHSVGFTPAATGGTGTGTC-----LQIOTSKPL 3711
Db 847 NVKRVYQESNENSAVQSIIPRPGIIVM--SEPTNODTSAPPTGMAAKPSRFQVYKSDAPI 90303
QY 372 VTSVAIVTVTVSLOPEKPVVSGTAVTSLPAPVTGEGSGAICLPYKPVVSPCWDHICK 431
Db 904 ATPISSTISSTATV--IPIVAAITPNTL-----TSEPIYOPITRAYIIT-----9433
QY 432 PVIQTGPOVQIKLAQBPV--LSQDAGIPTGSSSKOLFLEPHVVOVPGSGNEKQVTTISHS 488
Db 944 -----HLATSPSPVSHLSLSSNMS-PSATTHSNMSI-----GSTTSVPG 9800
QY 489 STLTIQCGQKTMVNNIITPISQPPPSILKQITLP-----GNKILSLQASPTQ 537
Db 981 RRTFVQPVQSAQESGISISTISIPHEPPIATISCPVPVSPVNVSTLMEVAPK 1037

```

```

RESULT 10
US-09-801-368-110
: Sequence 110. Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 110
LENGTH: 1075
TYPE: prt
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-110

```

Query Match	6.3%	Score 174;	DB 10;	Length 1075;
Best Local Similarity	21.8%;	Pred. No. 0.00072;		
Matches 139;	Conservative 80;	Mismatches 244;	Indels 174;	Gaps 29;
Oy	2 TLVYVAVAPVASAPKVVSSGRLPAPQIVAVKAPN-----TTTQFPANQLQPPGYLIKSN	56		

DT 01-NOV-1999 (TREMblrel. 12, last sequence update).
 DE 01-JUN-2002 (TREMblrel. 21, last annotation update).
 EG:5667.1 protein.
 OS Drosophila melanogaster (Fruit fly).
 GN Drosophila melanogaster.
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RAmannalides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RAGeorge R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RSutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RABrandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RAN K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RAdell J.F., Agbayani A., An H.-J., Andrews-Piankovich C., Baldwin D.,
 RBailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RBeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RBorkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RBurtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RCherry J.M., Carley S., Dahlke C., Davenport U.B., Davies P.,
 RAdePalos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RAdeson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RDurbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RFoster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,
 RGlodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RHarris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RHostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RJalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RKimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RLaško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RLiu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RMerkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RMount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RNeelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
 RPalazzo K., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RReinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RShue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RSoler E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RStyraskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RWang Z.-T., Wasserman D.A., Weinstock G.M., Weisenbach J.,
 RWilliams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RYe J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RZheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RAGibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "the genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cadieu E., Dreaano S., Lelaure V., Motlier S., Galibert F.,
 RT "Sequencing the distal x chromosome of Drosophila melanogaster."
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Benos P.
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AEO03421; AAF45644.1; -
 DR EMBL: AL01028; CA19845.2; -
 DR FLYBase: FBgn0025390; EG:5667.1.
 DR InterPro: IPR002557; Chitin_bind_PerA.
 DR InterPro: IPR002965; P_rich_extensn.
 DR Pfam: PF01607; CBM_14_2.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR SMART: SM00424; ChEBD2; 2.
 SQ SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;
 Query Match 7.8%; Score 214.5; DB 5; Length 1795;
 Best Local Similarity 20.2%; Pred. No. 9e-05;
 Matches 130; Conservative 86; Mismatches 254; Indels 173; Gaps 22;

QY 9 PVSADPKVSGPRLPAPQIVAVKAPNTTIOFPANLQI-----PGTVLIKNSG 58
 DB PTSSGKPTTPK-PSRTPTTKVTQTQITTTPLRSSSTETSTQPPPTTTPQPTT 758
 QY 59 PLMLVSPQOYTRATNTNITSRAVPANPQTVKICVTPNSSOLIKVAATVPKKLQI 118
 DB 759 TFLVTPKTSITTTTTERPITSSPK-PTTQKTSTAPNTT---KVAITQOETPT 812
 QY 119 GTT-----VTVTPKPS--VQSAVPTSVTVTPGKPLNT 152
 DB 813 QSTSTITTRKTTNNPEPTSEKPTSTTRKPTSTTSTKTSVASTKTTISSKPTTE 872
 QY 153 VTTKPSLSGASTPSNPNKAENSAVQINLSPTMLENVKCKNFIAMLIKLAGSSQ 212
 DB 873 KSTENPTNSVKTSLTSTQRA-----TSTSEPTKQNTITTPPKPTTLKTS 922
 QY 213 SPEKQNKILVEQLDLAK--TEAEFRKLVELKSSPQHLVFLKSVVALROLPN 270
 DB 923 TQEAFTSTQKSTVITTKKATSSPLTLTTEENTPPKP-----LRTTPT 970
 QY 271 SOSFIOCCVOOTSDMTATCTTVTSPVVTYTTSSOSEKSIIVGA--TAPR--TVS 326
 DB 971 TTS-----VTAITRITTTISESTSTSTQKRSKSTPTTTRTPVTVI 1017
 QY 327 VQTLNPLAGPVGAKGVVTLHSV-----PTATGTTA-GTGLQTSKPL 371
 DB 1018 VSTQNPTE--TTSKSTVITTPNPSPTQRTTTRQPTSTASTSIGTRITPTNP 1075
 QY 372 VTSVANVTYVTSLOPEK-----VSGTAVTSLPVTGEGSGAICLPVAVVS 424
 DB 1076 QNSTSTDLTYTRPCPDPTSDKNTTACTOELQVNLN-----LOSPOKQDF 1128
 QY 425 CWDHICKVIGT-----PVQIKIAPGVLSQAPGIPGSSSKQLFSIF 468
 DB 1129 THTRHTALTSRNTLGGQEPVDYMDAPSSAENSGCATTKAPTMCTILAAHLQKLF 1188
 QY 469 HVT-----QOPSGNEKQVTT-----ISHSTLTIQRCG 497
 DB 1189 HIISTTPSREHAPQRPSSQSSQSRGVYIAQMARHNLATSKPIAHSRLSIQOLA 1248
 QY 498 ---OKTMEVNTI-----PTSGPPASIIKQITLGNKIL 529
 DB 1249 STQKRSIPKTLVHTNTKEPDESEYDSEYDSEYDDEBNYL 1291
 RESULT 12
 Q8TDH7 PRELIMINARY; PRT; 1029 AA.
 AC Q8TDH7;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
 DE Intestinal membrane mucin MUC17 (Fragment).
 GN MUC17.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21845432; PubMed=11855812;
 RA Gum J.R., Jr., Crawley S.C., Hicks J.W., Szymkowski D.E., Kim Y.S.,
 RT "MUC17, a novel membrane-anchored mucin."
 RL Biochem. Biophys. Res. Commun. 291:466-475(2002).
 DR EMBL: AF430017; AAL89737.1; -
 FT NON_TER 1
 SQ SEQUENCE 1029 AA; 108320 MW; 5E1AD19E4BF948E3 CRC64;
 Query Match 7.8%; Score 214; DB 4; Length 1029;
 Best Local Similarity 21.9%; Pred. No. 4.9e-05;
 Matches 135; Conservative 76; Mismatches 259; Indels 146; Gaps 24;


```

OY 4 VPKAVASAPKVSQGRRLPAQIVAVKAPNTTIOFPALQLPFGVILKNSGMLMV 63
DB 137 VDSVTSTSTQSNSTP--TPREVITLPMSTPEVSTPLTI-MPVSTSTVISEAGTAST 193
OY 64 SPOQVTRAFETTSNITSRPAVPANPQVKICTVPPNSSOLIKKAAVTPVKLAOIGTVV 123
DB 194 LPVDISTPVITISQVSSSPVTPRG-TTMTPLWT-PSGEGSTLITMPVSTTVTSSEGSTLS 251
OY 124 TTVPKSSVQSVAVPSTSVVTPDGPPLNTVT-----TLKPSLGASSTPSNEPLKAENS 178
DB 252 T-----PSVATSTPTVSTSTEAISATLSDSTTMVSMPEISTIGTITLVSTPTVTRPES 307
OY 179 AAVQINLSPMLNKKCKKFLMLKLACSGSQSPMGQNVKLVQLDAIEAEFT 228
DB 308 TPTST--PSV-----YTSMTTASGSSSP-----TTEGT 337
OY 239 RKLVELKSSPOHVLPEFKKSVVALROLLPNSQSFIOQVQOTSSDMVATGTTVTS 298
DB 338 TTPMSTTSEKSTLLTVL---ISPIVMSPSASSTLSTPGDTSTPLSTAGSFTIP 394
OY 299 PVVTTVSSSQSEKSIIVSGATAPRTVSQVQTLNPLAGPVGAKAGVTLMSVGPATAGT 358
DB 395 AEVTTIRISTERS-----TPLTLLVST-----TL---PTSPFGAS 429
OY 359 TAGTGILQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTSLPAVTPGETSGAICLPV 418
DB 430 IASTPLDSTSTETPS---TDTASTPTLPV---ATTISVITIEGSGTTFIPST 480
OY 419 KPVVSEFCMD-----HICKPVI-----GTPVOIKLA--- 443
DB 481 -PVTSTADVFPATGAVSPVITSFELNPNSTSSSTSTTSFTKKEFTTPANTAPLR 539
OY 444 -----QPGVLSQAPGICPTGSSSKOLFLEHVQO-----PSGNEKQVTTISHST 490
DB 540 YVTMSTAPSTPRTISGCTASATLSTPHSTSVTPRVPSSSSPSTTSHTP 599
OY 491 LTIQKCGKTMVNTIIPTSQFPPA-SILKQTLPGKILSL-----QASPTQKNRIKEN 544
DB 600 PTPPPHSTPTTTSASSTVNEAVTMTTRKPKSTRTTSPTVTTTAVPT-NTTIKSN 658
OY 545 VMS-----CFRD 551
DB 659 PTSTPTVPRTTCFGD 674

RESULT 13
P91365 PRELIMINARY; PRT; 2232 AA.
AC P91365;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
GN K06A9.1 protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodietinae; Caenorhabditis.
OK NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RA Giesel C.; Gattling S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC OF THE A FORM.
CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
DR EMBL; U80846; AAC70889.1; -
DR EMBL; U80846; AAC70890.1; -
KM Alternative splicing.
FT VARSPLIC 842 866
FT VARSPLIC 870 1051
FT VARSPLIC 870 1051
CGSTSPGISTSEEMTSQSTQTPGSTGTVQPSVSDST

```

```

FT SSGSTVVGSTEGSSSPRISTSONTPSTSSGSGMSTQIPQ
FT SSGSTSPVESSTSGATSSSGSPGTTLSISPSPSSSTIGS
FT SSGSTSPVESSTISQSGTE -> KEIDOTAINITKYFENAL
FT LVASKLNNEISILGIDNFEGYSAGLNDHQYPPIDYNGKIS
FT VPPIDGTDIDDLDKVDKSLATADMPVPADONCMFTI
FT SAPEDEGCTTKSTKSTVYFVYGVYGVGAKISIPRLSDIK
FT NIVITNNIMNDRADASAVSKLLELPLFA (IN ISOFORM
FT B).
SQ SEQUENCE 2232 AA; 213840 MM; 08D69FA638E14CC8 CRC64;

Query Match 7.7%; Score 212.5; DB 5; Length 2232;
Best Local Similarity 21.9%; Pred. No. 0.00015;
Matches 138; Conservative 73; Mismatches 247; Indels 173; Gaps 24;

OY 26 QIVAVKAPNTTIOFPANQLPFGTVILKNSGMLVSPQVTRAFETTSNITSRPAVP 85
DB 801 QSVSTNPGSTVTR-----PSTVSGSTSSGTVTVGSTEASTSSSVASSAPASTS 852
OY 86 ANPQVVKICTVPPNSSOLIKKAAVTPVKLAOIGTVVTPRPSVQSVAVPSTSVTV 145
DB 853 QNPN-----PSTSS-----GSSMITQSPYPS-QSTSPVESSTPS 886
OY 146 PGKPLNTVTLKPS-----SIASSPNSN-----EPNL 173
DB 887 PGSPGTTLTSTSPSPSOSTTIGTQGSTPGCISTSEMTSQSTQTPGSTGTVQPT 946
OY 174 KAENSAA--VOINLS-----PTMLENVKCKNFKLMLKLACSGS---OSPEMGON 219
DB 947 VSDSTSSGTVTVGSTEGSSSPRISTSONTPSTSS-----SSSMSTQTPQSSQS 996
OY 220 VKRLVEQLDAIEAEFTKRLVELKSSPOPHL-----VPL 257
DB 997 TSP-VESSTSGANSSSGSPGTTLTLSISPSPSSSTIGSSGSTSPVSTISOGSTETPGS 1055
OY 258 KRSVALRQL-----PN-SQFIQCVQOTSSDMVIA 289
DB 1056 TGSVTVTPSTVSGSASGSTATMGSTASSTSGSSSTSPSPSOSTBSSTGATSSPGSSC 1115
OY 290 TCTTVTTSFVVTTVSSSOSEKSIIVSGATAPRTVSQVQTLNPLAGPVGAKAGVTLHSV 349
DB 1116 TLTLSISPSPSSSTIGSSGSTSPVSTSGDWTSGSTQIP-----GSTGTVQPT 1170
OY 350 GPTAAT-----GCTTAGTGLQTSKPLVTSVANTVTVSLOPEKPVVSGTAV---TL 388
DB 1171 GSGSTSGEITSGSGTQTPRSSLSTSPASTSTQOSVSTNS--PGSTVQPTSVRGSTS 1228
OY 399 SLPAVTFGETSG-----AALCLPSYKPVVSEFCMD-HICKPVIAGPVQIKLAOPGVLS 450
DB 1229 SGSTVTTTGSTEGSTSGSSSATSLSSSPVPSTISQSPNBTSGSSTPTPPSPGSTSPVVS 1288
OY 451 QPAGIPTGSSSKOLF-SLFHVQOP---SGNEKQVTTISHSTLTIQKCGKTMVNT-T 505
DB 1289 TTGEMTSHSGTQTPRTIGSTVQPTSVSGSNSGSGTVTIGSBSASTSGSFTSPSSIS 1348
OY 506 IIPTSQFPPASILKQTLPGKILSLQASPT 536
DB 1349 PVTSSPPTSTTFAAST-SGSTISDVSSVST 1378

RESULT 14
O9VGC9 PRELIMINARY; PRT; 842 AA.
AC O9VGC9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
GN CPN protein.
GN CPN OR CG4795.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_TaxID=7227;

```



```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Berens P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA Borokya D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Sytkus R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissensbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gdbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003694; AAF54755.1; -.
DR FLYbase; FBgn0010218; Cpn.
DR InterPro; IPR002965; P_flyc-extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
SQ SEQUENCE 842 AA; 82242 MW; D71E531327EF8501 CRC64;

Query Match 7.7%; Score 212; DB 5; Length 842;
Best Local Similarity 22.1%; Pred. No. 5.2e-05;
Matches 125; Conservative 81; Mismatches 236; Indels 124; Gaps 24;

OY 1 GILVTKV-APVSAP-----PKVSSGRLPAP-----QIVAVKAPNTTIIOPANQ 45
DB 4 GTTSPVASAPAAVPTTSAVAAPVOVSPAAPAAPIAVTPAAPPTLASVOAPAVT 63
OY 46 LPPTGLIKNSGPMILVSPQOTVTRAEITSNITSRPAVPANPQTVKICTVPNS--SSOLI 104
DB 64 IPAPAPIAASVAPASVAP--PVVAAPTPP-----AASPVSTPPVANAQIIVANASAVA 116
OY 105 KKVAVT-----PVKKLAOIGT--VTTVPKSSVQSAVPTSVTTVPKPLNTVTTLK 157
DB 117 PVVAATPPVAPAPIVAAPIATPPVAAAPPAATPVVSP--VIATPPVVPANTTV--- 171
OY 158 PSSLGASTPSENPILKAENSAVOINISPTMLE-----NKKCKKNFLAMLI 204
DB 172 PVAAPVAAVPAAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 228
OY 205 KLASGSGSPKMGQNVKLVLEQIDAKIEAEFEFRKLYVELKSSPOPHLVFLKSSVAL 264
DB 229 -----PEVSVAATKPLAAAEPPVVAAPATEIPVVAAPAAASHVSAVPAVERAAYA- 278
OY 265 ROLLPNOSGFIQOCVOQISSDMVIACTTTVTTSPPVTTTVSSSSEKSIIVSGATAPRT 324

```

```

DB 279 ----PVASAS-----TEPPVAAATLTAPET--PALAPVVAESQVAAVTVAATPPTP-A 324
OY 325 VSVOTLNPAGVGAAGVAVTSHVGPAAAGCTAGTGLDTSPLVTSVNTVTVSL 384
DB 325 PEPETIAP-----PVAAETPEVASAVAAETTPVPPVA-----AE 360
OY 385 OPEKPVSGTAVTSLPAVTEGETSGAAICLPSSVPPVS-----FCMDHICK 431
DB 361 SIPAPVVAATTPPARVL-AVTDPDVASAV--PELPPVIAPSPPVPSVAETPVDLAPVLP 417
OY 432 PYGTTPVOIKIAQPPPVLSQAPGIPFGSSSKOLELPHVVOQSGCNEKOVTTISHSTL 491
DB 418 PVAAEPVPAVVAEETPEPPAPASAVTTIAALDIPVAPVIAAPSDAPAE-----APSA 472
OY 492 TQKCGOKTMPVNTTIPTSQPPASI 517
DB 473 PT-----VSTPPTTASVPTTAPPAAV 494

RESULT 15
OY 095045 PRELIMINARY; PRT; 864 AA.
AC 095045;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE GH08002P.
GN Cpn OR CG4795.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyridae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Brockstein P., Hong L., Agbayan A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Flise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Paclab J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celiker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY058315; AALJ3544.1; -.
DR FLYbase; FBgn0010218; Cpn.
DR InterPro; IPR002965; P_flyc-extensn.
SQ SEQUENCE 864 AA; 84565 MW; E2B96CFB0CB9802 CRC64;

Query Match 7.7%; Score 212; DB 5; Length 864;
Best Local Similarity 22.1%; Pred. No. 5.3e-05;
Matches 125; Conservative 81; Mismatches 236; Indels 124; Gaps 24;

OY 1 GILVTKV-APVSAP-----PKVSSGRLPAP-----QIVAVKAPNTTIIOPANQ 45
DB 4 GTTSPVASAPAAVPTTSAVAAPVOVSPAAPAAPIAVTPAAPPTLASVOAPAVT 63
OY 46 LPPTGLIKNSGPMILVSPQOTVTRAEITSNITSRPAVPANPQTVKICTVPNS--SSOLI 104
DB 64 IPAPAPIAASVAPASVAP--PVVAAPTPP-----AASPVSTPPVANAQIIVANASAVA 116
OY 105 KKVAVT-----PVKKLAOIGT--VTTVPKSSVQSAVPTSVTTVPKPLNTVTTLK 157
DB 117 PVVAATPPVAPAPIVAAPIATPPVAAAPPAATPVVSP--VIATPPVVPANTTV--- 171
OY 158 PSSLGASTPSENPILKAENSAVOINISPTMLE-----NKKCKKNFLAMLI 204
DB 172 PVAAPVAAVPAAPVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 228
OY 205 KLASGSGSPKMGQNVKLVLEQIDAKIEAEFEFRKLYVELKSSPOPHLVFLKSSVAL 264
DB 229 -----PEVSVAATKPLAAAEPPVVAAPATEIPVVAAPAAASHVSAVPAVERAAYA- 278
OY 265 ROLLPNOSGFIQOCVOQISSDMVIACTTTVTTSPPVTTTVSSSSEKSIIVSGATAPRT 324

```

```

Dh 279 ----PV$AS-----TEPPVAATILTTAET- PALAPVAESQVAAANTVANTPTTP- A 324
Qy 325 V$VQTLNPLAOPVAKKAGVUTLH$VGPPLAATGTTAGTGLQTSKPLYTSVANTVTTVSL 364
Dh 325 PEPETIAF-----PVVAETPEVASVAEAETTPVPVPEVA-----AE 360
Qy 385 QPEKRV$SGTAVTSLPAWTEGTS$GAICLPSVKPV$S-----FCMDHICK 431
Dh 361 S1$P$VAVTTEVP$NTL-ATVDPDVTAS$V--PELPEPIA$SPVPG$VAETPEV$DL$APVLP 417
Qy 432 PVIGTPVOIKI$AQ$PVL$Q$P$AGIPTGSS$SKOLE$FL$HVQ$Q$B$GNEKOVTTISH$STL 491
Dh 418 PVAAEP$P$VAAV$EEETPE$P$AS$P$VITIA$LDIPE$VAPVIA$ASD$PAE-----AP$AA$ 472
Qy 492 TIQKQ$KTM$PVNTIITP$Q$PP$SI 517
Dh 473 P1-----V$TPPTTASV$ETIAP$P$AAV 494

```

Search completed: February 16, 2003, 21:58:53
Job time : 50.77 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 ; Search time 0.766595 Seconds
(without alignments)
1738.213 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 AQPGLVLSQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	53	100.0	801	19	AAW31494	Human hTAFII105 pr
2	53	100.0	801	22	ABG09468	Novel human diago
3	53	100.0	852	21	AA157279	Transcription fact
4	44	83.1	136	22	AAU45315	Protonibacterium
5	43	81.1	71	22	AAU06483	Human foetal prote
6	42	79.2	348	22	ABG14007	Novel human diago
7	41	77.4	353	21	AAU04409	Arabidopsis thalia
8	41	77.4	362	22	AAU04409	Maize disease resi
9	41	77.4	365	21	AAU04409	Arabidopsis thalia
10	41	77.4	397	21	AAU04409	Arabidopsis thalia

11	41	77.4	592	20	AAW97359	A 6-Transmembrane
12	40	75.5	63	23	ABP11105	Human ORFX protein
13	40	75.5	113	22	ABB03730	Human musculoskele
14	40	75.5	199	22	ABG20476	Novel human diago
15	39	73.6	24	22	AAE03611	Human hTAFII105
16	39	73.6	24	22	AAE03611	Human hTAFII105
17	39	73.6	46	22	AAE03611	Human hTAFII105
18	39	73.6	794	22	AAE03611	Human hTAFII105
19	39	73.6	794	22	AAE03611	Human hTAFII105
20	38	71.7	55	22	AAV12001	Human 5' EST seque
21	38	71.7	136	22	AAU00050	Human polypeptide
22	38	71.7	290	20	AAV48475	Human breast tumou
23	38	71.7	397	22	AAU95473	Human protein sequ
24	38	71.7	397	22	AAU95506	Human protein sequ
25	38	71.7	554	22	AAU95506	Human protein sequ
26	38	71.7	570	22	AAU95506	Human protein sequ
27	38	71.7	608	21	AAV1557	Human protein sequ
28	38	71.7	694	20	AAV17829	Human secreted pro
29	38	71.7	694	21	AAU01319	Human PRO354 prote
30	38	71.7	694	21	AAU01319	Human PRO354 polyp
31	38	71.7	694	22	AAU29053	Human PRO polypept
32	38	71.7	716	21	AAV1557	Human secreted pro
33	38	71.7	875	22	AAU88346	Human membrane or
34	38	71.7	942	21	AAV32169	Human growth-assoc
35	38	71.7	942	21	AAU95694	Human protein sequ
36	37	69.8	111	22	AAU95694	Human brain expres
37	37	69.8	111	22	AAU95694	Human bone marrow
38	37	69.8	111	22	AAU95694	Peptide #6084 enco
39	37	69.8	111	22	AAU95694	Human peptide enco
40	37	69.8	186	22	AAU95694	Fen-like protein.
41	37	69.8	194	20	AAU97465	Mouse TSI0423.3 ge
42	37	69.8	220	22	AAU94488	Human protein sequ
43	37	69.8	349	19	AAU83021	Mouse cysteine-rich
44	37	69.8	349	21	AAU83021	A murine Dickkopf
45	37	69.8	349	21	AAU92069	Murine DKR-3. Mus

ALIGNMENTS

RESULT 1
AAW31494 standard; Protein: 801 AA.
ID AAW31494:
AC AAW31494:
XX 28-APR-1998 (first entry)
DT Human hTAFII105 protein.
DE
XX
DE
XX
KW TAFII105, transcription factor, TFIID, transcriptional activation;
KW hTAFII105, transcription factor, TFIID, transcriptional activation;
KW antibodies; diagnosis; therapy; biopharmaceutical industry.
XX
XX Homo sapiens.
OS
XX
XX US5710025-A.
PN 20-JAN-1998.
PD 02-OCT-1996; 96US-0725012.
PR 02-OCT-1996; 96US-0725012.
PR (REGC) UNIV CALIFORNIA.
PI DiKstein R, Tjian R;
XX
XX WPI: 1998-109818/10.
XX N-PSDB: AAV02872.
XX
XX DNA encoding human tata-binding protein associated factor - for
XX producing recombinant protein

XX Claim 1: Col 17-22; 12pp; English.
 PS
 CC This cDNA sequence represents a human tata-binding protein associated
 CC factor, hTAFII105, isolated from Daudi cell nuclear extracts. Tightly
 CC associated subunits (TAF's) are components of the transcription factor
 CC TFIID and are thought to mediate transcriptional activation. This encoded
 CC protein may be produced recombinantly from transformed host cells or
 CC purified from human cells. hTAFII105 specific binding agents such as
 CC specific antibodies could be used for diagnosis (e.g. genetic
 CC hybridisation screens for hTAFII105 transcripts), therapy (e.g. gene
 CC therapy to modulate hTAFII105 gene expression) and in the
 CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
 CC cell specific activators or other transcriptional regulators).
 CC
 SQ Sequence 801 AA;
 Query Match 100.0%; Score 53; DB 19; Length 801;
 Best Local Similarity 100.0%; Pred. No. 0.94;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AOPGPVLSQP 10
 DB 443 AOPGPVLSQP 452
 RESULT 2
 ABG09468
 ID ABG09468 standard; Protein; 801 AA.
 AC
 XX ABG09468;
 AC
 XX 13-FEB-2002 (first entry)
 DT
 XX Novel human diagnostic protein #9459.
 DE
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 KM
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS73655.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PT
 XX Claim 20; SEQ ID No 39827; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 801 AA;
 Query Match 100.0%; Score 53; DB 22; Length 801;
 Best Local Similarity 100.0%; Pred. No. 0.94;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AOPGPVLSQP 10
 DB 443 AOPGPVLSQP 452
 RESULT 3
 AAY57279
 ID AAY57279 standard; Protein; 852 AA.
 AC
 XX AAY57279;
 AC
 XX 06-JUN-2000 (first entry)
 DT
 XX Transcription factor subunit TAFII105 polypeptide.
 DE
 XX TATA box-binding protein associated factor II 105; TAFII105; cancer;
 KW transcription factor; apoptosis; cytostatic; immunosuppressive;
 KM antiinflammatory; virucide; antibacterial.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 802
 FT /label= Glx
 FT /note= "encoded by TGA"
 FT Misc-difference 834
 FT /label= Glx
 FT /note= "encoded by TAA"
 XX
 PN WO200012699-A1.
 XX
 XX 09-MAR-2000.
 PD
 XX 25-AUG-1999; 99WO-IL00464.
 PF
 XX 27-AUG-1998; 98IL-0125971.
 PR
 XX (YEDA) YEDA RES & DEV CO LTD.
 PA
 XX Dikstein R, Yamit-hezi A;
 PI
 XX WPI: 2000-256640/22.
 DR N-PSDB; AAZ90465.
 XX
 PT Polypeptide encoding TATA box binding protein associated factor II 105
 PT useful for treating e.g. cancers and inducing apoptosis has a dominant
 PT negative effect on the normal biological activity of the binding
 PT protein
 PT
 XX Claim 7; Fig 2; 48pp; English.
 PS
 XX This represents a polypeptide comprising a (modified) fragment (I) of
 CC a TATA box-binding protein associated factor II 105 (TAFII105). A
 CC pharmaceutical composition comprising (I) or the polynucleotide or an

CC inhibitor or antagonist of (I) is useful for treating cancers and
CC inducing apoptosis in pathological cells. The composition is also useful
CC for treating autoimmune diseases, inflammatory processes and viral or
CC bacterial infections.

SO Sequence 852 AA;

Query Match 100.0%; Score 53; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSQP 10
|||
DB 443 AOPGPVLSQP 452

RESULT 4
AAU45315
ID AAU45315 standard; Protein; 136 AA.

AC AAU45315;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #6211.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PE 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71.

DR N-PSDB; AAS59525.

PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -

PS Example 1; SEQ ID NO 6510; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 136 AA;

Query Match 83.0%; Score 44; DB 22; Length 136;
Best Local Similarity 80.0%; Pred. No. 5.5;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AOPGPVLSQP 10
|||
DB 61 AOPGPVLSQP 70

RESULT 5
AAM06483
ID AAM06483 standard; Protein; 71 AA.

AC AAM06483;

DT 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 214.

XX Human: foetal protein; cytosolic; immunosuppressive; immunostimulant;
XX neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
XX gene therapy; antitense therapy; cancer; immune disorder;
XX nervous system disorder; osteoporosis; thrombolytic disorder;
XX nervous system disorder; inflammation.

OS Homo sapiens.

PN WO200155339-A2.

PD 02-AUG-2001.

PE 25-JAN-2001; 2001WO-US02723.

PR 25-JAN-2000; 2000US-0491404.

PR 15-SEP-2000; 2000US-0663870.

PR 06-NOV-2000; 2000US-0707351.

PA (HYSE-) HYSEQ INC.

PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
PI Liu C, Asundi V, Zhou P, Werhman T;

DR WPI; 2001-465571/50.

DR N-PSDB; AAH94158.

PT Novel fetal proteins useful for the treatment and diagnosis of diseases
PT associated with dysfunction of the protein e.g. cancers, immune
PT disorders, growth disorders, thrombolytic disorders, nervous system
PT disorders and inflammation -

PS Claim 10; Page 258; 715pp; English.

XX The invention relates to novel foetal polypeptides encoded by
XX polynucleotides comprising one of 477 sequences fully defined in the
XX specification. The foetal polynucleotides and polypeptides are
XX useful in the treatment and diagnosis of diseases such as cancers,
XX immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
XX disorders, nervous system disorders and inflammation. The present
XX sequence is a polypeptide encoded by a cDNA assembled using
XX an expressed sequence tag (EST) found to be expressed in human
XX foetal tissue cDNA libraries.

SO Sequence 71 AA;

Query Match 81.1%; Score 43; DB 22; Length 71;
Best Local Similarity 70.0%; Pred. No. 4.3;

Matches 7: Conservative 3: Mismatches 0: Indels 0: Gaps 0:

OY 1 AOPGPVLSOP 10
Db 40 SRPGPVLSOP 49

RESULT 6

ABG14007
ID ABG14007 standard; Protein; 348 AA.

AC ABG14007;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #13998.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.
DR N-PSDB; AAS78194.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID NO 44366; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 348 AA;

Query Match 79.2%; Score 42; DB 22; Length 348;
Best Local Similarity 80.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
Db 99 AOPGPVLSOP 108

RESULT 7

AAG29893
ID AAG29893 standard; Protein; 353 AA.

AC AAG29893;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 35642.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 17-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.4%; Score 41; DB 21; Length 353;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AOPCPVLS 8
|||||||
Db 154 AOPCPVLS 161

RESULT 8
AAU00449

ID AU00449 standard; Protein: 362 AA.
XX
AC AAU00449;
XX
DT 19-JUN-2001 (first entry)
XX
DE Maize disease resistance enhancing protein ZmPtl1-1.
XX
KW Maize; plant disease resistance; crop; soybean; sunflower; sorghum;
KM canola; wheat; alfalfa; cotton; rice; barley; millet; plant cell death;
KW herbicide resistance; ZmPtl1-1; Ptl1; Ptol.
XX
OS Zea mays.
XX
PN WO200118061-A2.
XX
PD 15-MAR-2001.
XX
PF 06-SEP-2000; 2000WO-US24403.
XX
PR 09-SEP-1999; 99US-0152988.
XX
PA (P10N-) PIONEER HI-BRED INT INC.
XX
PI Simmons CR;
XX
PT WPT: 2001-226742/23.
DR N-PSDB: AAS01019.
XX
DR Novel isolated maize disease resistance polynucleotide useful for
PT increasing resistance in a plant to disease, controlling cell death,
PT and conferring resistance to herbicides -
XX
PS Claim 11; Page 89-90; 90pp; English.
XX
XX The present sequence representing maize ZmPtl1-1 protein is 1 of 7
CC novel disease resistance proteins (AAU00443-AAU00449). The ZmPtl1-1 is
CC a Zea mays gene homologue to the plant disease resistance genes Ptl1
CC and Ptol in tomato. These novel maize disease resistance polynucleotides
CC and polypeptides are useful for enhancing disease resistance in crops
CC and transgenic plants including maize, soybean, sunflower, sorghum,
CC canola, wheat, alfalfa, cotton, rice, barley or millet. The level of
CC disease resistance protein is increased by transforming a plant cell
CC with a recombinant expression cassette which comprises the disease
CC resistance polynucleotide operably linked to a promoter, or by culturing
CC the plant cell under plant growing conditions to produce a regenerated
CC plant, or by inducing expression of the polynucleotide to modulate the
CC disease resistance protein in a plant. The polynucleotides encoding the
CC disease resistance proteins are useful for increasing resistance in a
CC plant to disease, controlling cell death, and conferring resistance to
CC herbicides. They are useful as probes or amplification primers in the
CC detection, quantification, or isolation of gene transcripts. They can be
CC used for recombinant expression or isolation of gene transcripts, as
CC immunogens in the preparation and/or screening of antibodies, and in
CC sense or antisense suppression of the polynucleotide in a host cell,
CC tissue or plant.
XX
SQ Sequence 362 AA;
XX
Query Match 77.4%; Score 41; DB 22; Length 362;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35641.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP103405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 25-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150864.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158359.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.4%; Score 41; DB 21; Length 365;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AOPCPVLS 8
Db 166 AOPCPVLS 173

RESULT 10
ID AAG29891
AAG29891 standard; Protein: 397 AA.

XX AAG29891;
XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 35640.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.
OS Arabidopsis thaliana.
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126762.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.4%; Score 41; DB 21; Length 397;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQP6PVLS 8
Db 198 AQP6PVLS 205

RESULT 11
AAW97359

ID AAW97359 standard; Protein; 592 AA.

AC AAW97359;

DT 12-MAY-1999 (first entry)

DE A 6-Transmembrane protein designated SDR2.

KW 6-Transmembrane protein; SDR2; cancer; inflammation; autoimmunity;
KW allergy; asthma; rheumatoid arthritis; CNS inflammation;
KW cerebellar degeneration; Alzheimer's disease; Parkinson's disease;
KW multiple sclerosis; amyotrophic lateral sclerosis; head injury damage;
KW neurological abnormality; septic shock; sepsis; stroke; osteoporosis;
KW osteoarthritis; ischaemia reperfusion injury; cardiovascular disease;
KW kidney disease; liver disease; ischaemic injury; myocardial infarction;
KW hypotension; hypertension; AIDS; myelodysplastic syndrome;
KW hematological abnormality; aplastic anaemia; male pattern baldness;

KW Infection.
XX
OS Homo sapiens.
XX
PN EP897979-A2.
XX
PD 24-FEB-1999.
XX
PF 06-AUG-1998; 98EP-0306292.
XX
PR 16-DEC-1997; 97US-0991813.
PR 12-AUG-1997; 97US-0055375.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Albion EF, Kikly KK;
XX WPI; 1999-134642/12.
DR N-PSDB; MAX15876.
XX
PT New 6-Transmembrane protein (SDR2) polypeptide and polynucleotide --
PT useful as diagnostic reagents and for prevention and treatment of
PT cancer, AIDS and microbial infections
XX
PS Claim 2; Page 6; 19pp; English.

CC The present sequence represents a 6-Transmembrane protein polypeptide
CC designated SDR2. SDR2 polypeptides and polynucleotides are useful for
CC diagnosing susceptibility to diseases by detecting mutations in the SDR2
CC gene, and can diagnose diseases associated with SDR2 protein imbalance by
CC determining SDR2 polypeptide expression levels. SDR2 polypeptides can be
CC used to screen for agonists and antagonists and to produce antibodies.
CC Diseases diagnosed, prevented and treated include: cancer, inflammation,
CC autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,
CC cerebellar degeneration, Alzheimer's disease, Parkinson's disease,
CC multiple sclerosis, amyotrophic lateral sclerosis, head injury damage,
CC and other neurological abnormalities, septic shock, sepsis, stroke,
CC osteoporosis, osteoarthritis, ischaemia reperfusion injury,
CC cardiovascular disease, kidney disease, liver disease, ischaemic injury,
CC myocardial infarction, hypotension, hypertension, AIDS, myelodysplastic
CC syndromes and other hematological abnormalities, aplastic anaemia, male
CC pattern baldness, and bacterial, fungal, protozoan and viral infections.
XX
SQ Sequence 592 AA;

Query Match 77.4%; Score 41; DB 20; Length 592;
Best Local Similarity 75.0%; Pred. No. 74;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 PGVYLSQP 10
Db 158 PGVYLSQP 165

RESULT 12
ABP11105

ID ABP11105 standard; Protein; 63 AA.

AC ABP11105;

DT 24-JUN-2002 (first entry)

DE Human ORF protein sequence SRQ ID NO:22192.

KW Human; open reading frame; ORF; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.

PA (HYSE-) HYSEQ INC.
 XX Drmanac RT, Jlu C, Tang YT;
 XX WPI: 2001-639362/73.
 DR N-PSDB: AAS84663.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20: SEQ ID No 50835; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG3037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SO Sequence 199 AA:
 QY
 DB 173 AOPGPVLSOP 10
 173 AOPGPVGSNP 182
 DB
 RESULT 15
 AAE03611
 ID AAE03611 standard; peptide; 24 AA.
 XX
 AC AAE03611;
 XX
 DT 07-AUG-2001 (first entry)
 XX
 DE Human AZAD protein leucine-rich repeat sequence #7.
 XX
 XX Human; leucine-rich repeat; LRR; AZAD; neurodegenerative disorder;
 KW CNS disorder; central nervous system disorder; prostate disorder;
 KW prostatic; benign prostatic hyperplasia; adenocarcinoma;
 KW prostate cancer; genitourinary system carcinoma; testicular tumour;
 KW Alzheimer's disease; dementia; Parkinson's disease; multiple sclerosis;
 KW amyotrophic lateral sclerosis; epilepsy; psychiatric disorder; mania;
 KW depression; schizophrenia; anxiety; phobic disorder; learning disorder;
 KW memory disorder; amnesia; migraine; protein-protein interaction;
 KW cellular activity; neoplastic transformation; drug screening;
 KW forensic identification; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200142286-A2.
 XX
 PD 14-JUN-2001.

XX
 PF 07-DEC-2000; 2000WO-0533140.
 XX
 PR 08-DEC-1999; 99US-0456592.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Rhodadoust MM;
 XX
 DR WPI: 2001-381633/40.
 XX
 PT New leucine-rich repeat-containing protein and nucleic acid molecules
 PT for diagnosing, treating neural disorders, such as neurodegenerative
 PT disorders, such as Alzheimer's disease, dementia, epilepsy and prostate
 PT cancer -
 XX
 PS Example 1; Fig 4; 133pp; English.
 XX
 CC The present sequence is a leucine-rich repeat of
 CC human AZAD protein which is a multiple leucine-rich repeat-containing
 CC secreted protein. AZAD polypeptide is useful for
 CC identifying a compound which modulates its activity and binds to it.
 CC AZAD proteins and nucleic acid molecules are useful for treating and
 CC diagnosing AZAD-mediated or related disorders, which includes a neural
 CC disorder (e.g. neurodegenerative disorders, including CNS disorders)
 CC and a prostate disorder, e.g. prostaticitis, benign prostatic hyperplasia
 CC and cancer (e.g. adenocarcinoma, prostate cancer, genitourinary system
 CC carcinomas and testicular tumour). The neurodegenerative disorders
 CC include Alzheimer's disease, dementias related to Alzheimer's disease,
 CC Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis,
 CC epilepsy, psychiatric disorders e.g. depression, schizophrenia
 CC disorders, mania, anxiety or phobic disorders, learning or memory
 CC disorders e.g. amnesia or age-related memory loss, and neurological
 CC disorders e.g. migraine. AZAD molecules are useful as markers of
 CC disorders or disease states, as markers for precursors of disease state,
 CC for predisposition of disease state or as markers of drug activity or
 CC pharmacogenomic profile of a subject. The AZAD polypeptide and
 CC polynucleotide are capable of modulating protein-protein interaction,
 CC e.g. by interacting with an extracellular component, thereby modulating
 CC cellular activities, including attachment, adhesion, migration,
 CC patterning, growth and/or differentiation of a cell. AZAD proteins
 CC regulate embryonic development and differentiation, tissue maintenance
 CC and function, pathological conditions, e.g. neuronal degeneration,
 CC neoplastic transformation and tumour progression. AZAD proteins are
 CC useful as immunogens to raise anti-AZAD antibodies which are useful to
 CC detect and isolate AZAD proteins and modulate its activity. AZAD
 CC proteins are useful to screen for naturally occurring AZAD substrates
 CC and to screen for drugs or compounds which modulate AZAD activity.
 CC AZAD nucleic acid fragments are useful as primers or hybridisation
 CC probes for the detection of AZAD-encoding nucleic acids. cDNA encoding
 CC the AZAD protein are useful in gene therapy. AZAD sequences are also
 CC useful to map their respective genes on a chromosome, for tissue typing
 CC and in forensic identification of a biological sample.
 CC
 SO Sequence 24 AA:
 QY
 DB 14 ALRPPVLSO 22
 1 AOPGPVLSO 9
 1 AOPGPVLSO 22
 DB
 Query Match 73.6%; Score 39; DB 22; Length 24;
 Best Local Similarity 88.9%; Pred. No. 7;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Search completed: February 16, 2003, 21:55:08
 Job time : 2.7666 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 0.256959 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53
Sequence: 1 AOPGVLSP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/55_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	801	1	US-08-725-012-2
2	41	77.4	592	3	US-08-991-813-2
3	38	71.7	942	3	US-09-074-579-1
4	38	71.7	942	4	US-09-388-774-1
5	37	69.8	186	4	US-09-325-932A-72
6	37	69.8	349	4	US-09-161-241-8
7	37	69.8	645	4	US-08-791-115B-6
8	37	69.8	756	2	US-08-887-518-4
9	37	69.8	756	2	US-09-023-321-4
10	37	69.8	756	2	US-08-890-853-2
11	37	69.8	756	2	US-09-032-475-4
12	37	69.8	756	2	US-09-099-125A-2
13	37	69.8	756	2	US-09-099-124A-2
14	37	69.8	756	4	US-09-032-476-2
15	37	69.8	756	4	US-08-890-854-2
16	37	69.8	756	4	US-09-023-324-2
17	37	69.8	756	4	US-09-168-629-15
18	37	69.8	756	4	US-08-910-820-9
19	36	67.9	362	2	US-09-055-097-7
20	36	67.9	404	3	US-08-459-953A-7
21	35	66.0	115	4	US-08-630-915A-10
22	35	66.0	350	4	US-09-161-241-9
23	34	64.2	249	4	US-09-345-468-9
24	34	64.2	249	4	US-09-414-453A-9
25	34	64.2	307	4	US-09-134-001C-3238
26	34	64.2	319	4	US-09-345-468-5
27	34	64.2	319	4	US-09-414-453A-5

28	34	64.2	328	2	US-08-651-818A-2	Sequence 2, Appl1
29	34	64.2	328	4	US-09-184-826-2	Sequence 2, Appl1
30	34	64.2	339	4	US-09-345-468-3	Sequence 3, Appl1
31	34	64.2	339	4	US-09-414-453A-3	Sequence 3, Appl1
32	34	64.2	384	4	US-08-221-750A-11	Sequence 11, Appl1
33	34	64.2	565	4	US-08-906-156A-12	Sequence 12, Appl1
34	34	64.2	742	4	US-08-791-115B-5	Sequence 5, Appl1
35	34	64.2	830	4	US-09-562-737-38	Sequence 38, Appl1
36	34	64.2	2627	2	US-08-751-189-3	Sequence 3, Appl1
37	34	64.2	2627	2	US-09-060-836-3	Sequence 3, Appl1
38	34	64.2	2627	3	US-09-184-445-3	Sequence 3, Appl1
39	33	62.3	222	3	US-08-476-509B-40	Sequence 40, Appl1
40	33	62.3	176	2	US-08-619-542B-44	Sequence 44, Appl1
41	33	62.3	180	4	US-09-133-341-12	Sequence 12, Appl1
42	33	62.3	444	4	US-09-129-668-8	Sequence 8, Appl1
43	33	62.3	464	2	US-08-969-630-5	Sequence 5, Appl1
44	33	62.3	580	3	US-08-906-865-1	Sequence 1, Appl1
45	33	62.3	580	4	US-09-129-668-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-725-012-2
Sequence 2, Application US/08725012
Patent No. 5710025
GENERAL INFORMATION:
APPLICANT: Dikstein, Rivka
APPLICANT: Tjian, Robert
TITLE OF INVENTION: B-Cell Specific Transcription Factor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-012-2
Query Match 100.0%; Score 53; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGVLSP 10
|||||||
Db 443 AOPGVLSP 452
RESULT 2

US-08-991-813-2
; Sequence 2, Application US/08991813
; Patent No. 6050579
; GENERAL INFORMATION:
; APPLICANT: ALBONE, EARL
; APPLICANT: KIKLY, KRISTINE
; TITLE OF INVENTION: HUMAN SDR2 CDNA CLONE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,813
; FILING DATE: 16-DEC-1997
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/055,375
; FILING DATE: 12-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 592 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-991-813-2

Query Match 77.4%; Score 41; DB 3; Length 592;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 QGPVLSQP 10
|||||

DB 158 QGPVLSQP 165

RESULT 3
US-09-074-579-1
; Sequence 1, Application US/09074579
; Patent No. 6001596
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,579
FILING DATE: HEREWITH
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0505 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 942 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: UTRSN0702
CLONE: 688183
US-09-074-579-1

Query Match 71.7%; Score 38; DB 3; Length 942;
Best Local Similarity 66.7%; Pred. No. 11e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGPVLSQP 10
|||||

DB 658 QGPVLSQP 666

RESULT 4
US-09-388-774-1
; Sequence 1, Application US/09388774
; Patent No. 6228991
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,774
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 09/074,579
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0505 US
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 942 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: UTRSN0702
CLONE: 688183
US-09-388-774-1

Query Match 71.7%; Score 38; DB 4; Length 942;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGVLSOP 10
DB 658 QPGPLKKP 666

RESULT 5
US-09-325-932A-72
Sequence 72, Application US/09325932A
Patent No. 6451604
GENERAL INFORMATION:
APPLICANT: Flihn, Barry
APPLICANT: Lasham, Annette
TITLE OF INVENTION: Compositions affecting programmed cell
FILE REFERENCE: 1022
CURRENT APPLICATION NUMBER: US/09/325,932A
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 206
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 72
LENGTH: 186
TYPE: PRT
ORGANISM: Pinus radiata
US-09-325-932A-72

Query Match 69.8%; Score 37; DB 4; Length 186;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQPGLVL 7
DB 166 AQPGLVL 172

RESULT 6
US-09-161-241-8
Sequence 8, Application US/09161241
Patent No. 6344541
GENERAL INFORMATION:
APPLICANT: Bass, Michael B
APPLICANT: Sullivan, John K
APPLICANT: Theill, Lars E
APPLICANT: Wang, Daquan
TITLE OF INVENTION: NOVEL DKR POLYPEPTIDES
FILE REFERENCE: A-548
CURRENT APPLICATION NUMBER: US/09/161,241
CURRENT FILING DATE: 1998-09-25
NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patent Ver. 2.0
SEQ ID NO: 8
LENGTH: 349
TYPE: PRT
ORGANISM: Mouse
US-09-161-241-8

Query Match 69.8%; Score 37; DB 4; Length 349;
Best Local Similarity 60.0%; Pred. No. 56;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AQPGLVL 10
DB 32 AEPGLNLP 41

RESULT 7
US-08-791-115B-6
Sequence 6, Application US/08791115B
Patent No. 6262242
GENERAL INFORMATION:
APPLICANT: Steck, Peter
APPLICANT: Pershouse, Mark A.
APPLICANT: Jasser, Samir
APPLICANT: Yung, W.K. Alfred
TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
STREET: 565 Thirteenth Street, N.W., Suite 701-E
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 22204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/791,115B
FILING DATE: 30-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 38,957
REFERENCE/DOCKET NUMBER: 2318-134.A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-683-6040
TELEFAX: 202-683-7031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 645 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-791-115B-6

Query Match 69.8%; Score 37; DB 4; Length 645;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PGPVLSOP 10
DB 126 PGPVLSLP 133

RESULT 8
US-08-887-518-4
Sequence 4, Application US/08887518
Patent No. 5843721
GENERAL INFORMATION:
APPLICANT: Rothe, Mike
APPLICANT: Wu, Lin
TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200

CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,518
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-887-518-4

Query Match
Best Local Similarity 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
:||||:||||
Db 682 SGPQQLMSQP 691

RESULT 9
US-09-023-321-4
Sequence 4, Application US/09023321
Patent No. 5844073
GENERAL INFORMATION:
APPLICANT: Rolhe, Mike
TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,321
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/887,518
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342

INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-321-4

Query Match
Best Local Similarity 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
:||||:||||
Db 682 SGPQQLMSQP 691

RESULT 10
US-08-890-853-2
Sequence 2, Application US/08890853
Patent No. 5851812
GENERAL INFORMATION:
APPLICANT: Goeddel, David V.
TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/890,853
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-890-853-2

Query Match
Best Local Similarity 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
:||||:||||
Db 682 SGPQQLMSQP 691

RESULT 11
US-09-032-475-4
Sequence 4, Application US/09032475
Patent No. 5854003
GENERAL INFORMATION:

APPLICANT: Rothe, Mike
APPLICANT: Wu, Lin
TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,475
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/887,518
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-032-475-4
Query Match 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 AOPGPVLSOP 10
Db 682 SQPGQLMSOP 691
RESULT 12
US-09-099-125A-2
Sequence 2, Application US/09099125A
Patent No. 5916760
GENERAL INFORMATION:
APPLICANT: Goeddel, David V.
APPLICANT: Moronitz, John
TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,125A
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/890,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-099-125A-2
Query Match 69.8%; Score 37; DB 2; Length 756;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 AOPGPVLSOP 10
Db 682 SQPGQLMSOP 691
RESULT 13
US-09-099-124A-2
Sequence 2, Application US/09099124A
Patent No. 5939302
GENERAL INFORMATION:
APPLICANT: Goeddel, David V.
APPLICANT: Moronitz, John
TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,124A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/890,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: T97-006-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 756 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-099-124A-2
Query Match 69.8%; Score 37; DB 2; Length 756;

Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGYLSOP 10
:||||:||||
Db 682 SOPGQLMSOP 691

RESULT 14
US-09-032-476-2

; Sequence 2, Application US/09032476
; Patent No. 6235492

; GENERAL INFORMATION:

; APPLICANT: Rothe, Mike
; APPLICANT: Cao, Zhaoan

; APPLICANT: R guler, Catherine

; TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 268 BUSH STREET, SUITE 3200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/032,476

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/890,854

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: T97-006-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 343-4341

; TELEFAX: (415) 343-4342

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 756 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-09-032-476-2

Query Match 69.8%; Score 37; DB 4; Length 756;

Best Local Similarity 60.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGYLSOP 10
:||||:||||
Db 682 SOPGQLMSOP 691

RESULT 15
US-08-890-854-2

; Sequence 2, Application US/08890854
; Patent No. 6235512

; GENERAL INFORMATION:

; APPLICANT: Rothe, Mike

; APPLICANT: Cao, Zhaoan

; APPLICANT: R guler, Catherine

; TITLE OF INVENTION: IKK- Proteins, Nucleic Acids and Methods

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/890,854

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: T97-006-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 343-4341

TELEFAX: (415) 343-4342

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 756 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-890-854-2

Query Match 69.8%; Score 37; DB 4; Length 756;

Best Local Similarity 60.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGYLSOP 10
:||||:||||
Db 682 SOPGQLMSOP 691

Search completed: February 16, 2003, 22:02:59
Job time: 2.25696 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 0.197002 Seconds
(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452
Perfect score: 53
Sequence: 1 AQPGLVSLQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	75.5	113	10 US-09-764-877-1677	Sequence 1677, Ap
2	39	73.6	24	10 US-09-789-404-14	Sequence 14, Appl
3	39	73.6	794	10 US-09-789-404-2	Sequence 2, Appl
4	38	71.7	694	9 US-09-944-413-55	Sequence 55, Appl
5	38	71.7	694	9 US-09-944-403-55	Sequence 55, Appl
6	38	71.7	694	9 US-09-944-896-55	Sequence 55, Appl
7	38	71.7	694	9 US-09-944-944-55	Sequence 55, Appl
8	38	71.7	694	9 US-09-944-929-55	Sequence 55, Appl
9	38	71.7	694	9 US-09-944-929-55	Sequence 55, Appl
10	38	71.7	694	9 US-10-174-590-60	Sequence 60, Appl
11	38	71.7	694	9 US-10-176-558-60	Sequence 60, Appl
12	38	71.7	694	9 US-10-175-737-60	Sequence 60, Appl
13	38	71.7	694	9 US-10-173-706-60	Sequence 60, Appl
14	38	71.7	694	9 US-10-175-738-60	Sequence 60, Appl
15	38	71.7	694	9 US-10-175-752-60	Sequence 60, Appl
16	38	71.7	694	9 US-10-176-482-60	Sequence 60, Appl
17	38	71.7	694	9 US-10-176-757-60	Sequence 60, Appl
18	38	71.7	694	9 US-10-176-913-60	Sequence 60, Appl
19	38	71.7	694	9 US-10-180-552-60	Sequence 60, Appl

20	38	71.7	694	9 US-10-180-557-60	Sequence 60, Appl
21	38	71.7	694	9 US-10-173-700-60	Sequence 60, Appl
22	38	71.7	694	9 US-10-174-572-60	Sequence 60, Appl
23	38	71.7	694	9 US-10-174-579-60	Sequence 60, Appl
24	38	71.7	694	9 US-10-174-582-60	Sequence 60, Appl
25	38	71.7	694	9 US-10-174-588-60	Sequence 60, Appl
26	38	71.7	694	9 US-10-175-739-60	Sequence 60, Appl
27	38	71.7	694	9 US-10-175-740-60	Sequence 60, Appl
28	38	71.7	694	9 US-10-175-743-60	Sequence 60, Appl
29	38	71.7	694	9 US-10-176-488-60	Sequence 60, Appl
30	38	71.7	694	9 US-10-176-492-60	Sequence 60, Appl
31	38	71.7	694	9 US-10-176-747-60	Sequence 60, Appl
32	38	71.7	694	9 US-10-176-750-60	Sequence 60, Appl
33	38	71.7	694	9 US-10-176-985-60	Sequence 60, Appl
34	38	71.7	694	9 US-10-176-987-60	Sequence 60, Appl
35	38	71.7	694	9 US-10-176-991-60	Sequence 60, Appl
36	38	71.7	694	9 US-10-176-992-60	Sequence 60, Appl
37	38	71.7	694	9 US-10-176-993-60	Sequence 60, Appl
38	38	71.7	694	9 US-10-184-658-60	Sequence 60, Appl
39	38	71.7	694	10 US-09-866-028-55	Sequence 55, Appl
40	38	71.7	694	10 US-09-944-449-55	Sequence 55, Appl
41	38	71.7	694	10 US-09-944-457-55	Sequence 55, Appl
42	38	71.7	694	10 US-09-944-862-55	Sequence 55, Appl
43	38	71.7	694	10 US-09-945-587-55	Sequence 55, Appl
44	38	71.7	694	10 US-09-945-015-55	Sequence 55, Appl
45	38	71.7	694	10 US-09-944-396-55	Sequence 55, Appl

ALIGNMENTS

RESULT 1
US-09-764-877-1677
; Sequence 1677, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; PRIOR APPLICATION DATA REMOVED - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1677
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (67)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-877-1677

Query Match 75.5% Score 40; DB 10; Length 113;
Best Local Similarity 77.8% Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QCPGLVSLQP 10
|||||
Db 102 QCPGLVSLQP 110

RESULT 2
US-09-789-404-14
; Sequence 14, Application US/09789404
; Patent No. US20020025554A1
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL LEUCINE RICH REPEAT-CONTAINING MOLECULES AND USES THEREOF

FILE REFERENCE: 10448/008001
CURRENT APPLICATION NUMBER: US/09/789,404
CURRENT FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 09/456,592
PRIOR FILING DATE: 1999-12-08
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 24
TYPE: PRT
ORGANISM: Homo sapiens
US-09-789-404-14

Query Match
Best Local Similarity 73.6%; Score 39; DB 10; Length 24;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQ 9
Db 14 ALPGPVLSQ 22

RESULT 3
US-09-789-404-2
Sequence 2, Application US/09789404
Patent No. US2002002554A1
GENERAL INFORMATION:
APPLICANT: Knodadoust, Mehnan
TITLE OF INVENTION: NOVEL LEUDINE RICH REPEAT-CONTAINING MOLECULES AND USES THEREFOR
FILE REFERENCE: 10448/008001
CURRENT APPLICATION NUMBER: US/09/789,404
CURRENT FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 09/456,592
PRIOR FILING DATE: 1999-12-08
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 794
TYPE: PRT
ORGANISM: Homo sapiens
US-09-789-404-2

Query Match
Best Local Similarity 73.6%; Score 39; DB 10; Length 794;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOPGPVLSQ 9
Db 253 ALPGPVLSQ 261

RESULT 4
US-09-944-413-55
Sequence 55, Application US/09944413
Patent No. US20020156004A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Baton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerltsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kljavin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944,413
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: No. US20020156004A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020156004A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/33565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042

PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 55
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Sapien
US-09-944-413-55

Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGVLSOP 10
DB 658 QPGPLKRP 666

RESULT 5
US-09-944-403-55
Sequence 55, Application US/09944403
Patent No. US20020165143A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerltsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944,403
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997

PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: No. US20020165143A1, September 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020165143A1, September 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 55
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Sapien
US-09-944-403-55

Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGVLSOP 10
DB 658 QPGPLKRP 666

RESULT 6
US-09-944-896-55
Sequence 55, Application US/09944896
Patent No. US20020168715A1
GENERAL INFORMATION:

APPLICANT: Baker, Kevin
APPLICANT: Batstein, David
APPLICANT: Batson, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerltzen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944,896
CURRENT FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: NO. US20020168715a1ember 30, 1999

PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: NO. US20020168715a1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 55
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Saplen
US-09-944-896-55
Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 QPGVLSQP 10
db 658 QPGVLKRP 666
RESULT 7
US-09-944-944-55
Sequence 55, Application US/099444944
Patent No. US20020173463A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Batstein, David
APPLICANT: Batson, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerltzen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944,944
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278

;; PRIOR FILING DATE: December 11, 1997
;; PRIOR APPLICATION NUMBER: 60/069,425
;; PRIOR FILING DATE: December 12, 1997
;; PRIOR APPLICATION NUMBER: 60/069,696
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,694
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,702
;; PRIOR FILING DATE: December 16, 1997
;; PRIOR APPLICATION NUMBER: 60/069,870
;; PRIOR FILING DATE: December 17, 1997
;; PRIOR APPLICATION NUMBER: 60/069,873
;; PRIOR FILING DATE: December 17, 1997
;; PRIOR APPLICATION NUMBER: 60/068,017
;; PRIOR FILING DATE: December 18, 1997
;; PRIOR APPLICATION NUMBER: 60/070,440
;; PRIOR FILING DATE: January 5, 1998
;; PRIOR APPLICATION NUMBER: 60/074,086
;; PRIOR FILING DATE: February 9, 1998
;; PRIOR APPLICATION NUMBER: 60/074,092
;; PRIOR FILING DATE: February 9, 1998
;; PRIOR APPLICATION NUMBER: 60/075,945
;; PRIOR FILING DATE: February 25, 1998
;; PRIOR APPLICATION NUMBER: 60/112,850
;; PRIOR FILING DATE: December 16, 1998
;; PRIOR APPLICATION NUMBER: 60/113,296
;; PRIOR FILING DATE: December 22, 1998
;; PRIOR APPLICATION NUMBER: 60/146,222
;; PRIOR FILING DATE: July 28, 1999
;; PRIOR APPLICATION NUMBER: PCT/US98/19330
;; PRIOR FILING DATE: September 15, 1998
;; PRIOR APPLICATION NUMBER: PCT/US98/25108
;; PRIOR FILING DATE: December 1, 1998
;; PRIOR APPLICATION NUMBER: 09/216,021
;; PRIOR FILING DATE: December 16, 1998
;; PRIOR APPLICATION NUMBER: 09/218,517
;; PRIOR FILING DATE: December 22, 1998
;; PRIOR APPLICATION NUMBER: 09/254,311
;; PRIOR FILING DATE: March 3, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/12252
;; PRIOR FILING DATE: June 22, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: September 15, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28409
;; PRIOR FILING DATE: NO. US20020173463A1eember 30, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: NO. US20020173463A1eember 30, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/28301
;; PRIOR FILING DATE: December 1, 1999
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: December 16, 1999
;; PRIOR APPLICATION NUMBER: PCT/US00/03565
;; PRIOR FILING DATE: February 11, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: February 22, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/05841
;; PRIOR FILING DATE: March 2, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/08439
;; PRIOR FILING DATE: March 30, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/14042
;; PRIOR FILING DATE: May 22, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/20710
;; PRIOR FILING DATE: July 28, 2000
;; PRIOR APPLICATION NUMBER: PCT/US00/32678
;; PRIOR FILING DATE: December 1, 2000
;; PRIOR APPLICATION NUMBER: PCT/US01/06520
;; PRIOR FILING DATE: February 28, 2001
;; NUMBER OF SEQ ID NOS: 120
;; SEQ ID NO 55
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-944-944-55

Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. NO. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
||||:|
Db 658 QPGPVLKRP 666

RESULT 8
US-09-944-907-55

;; Sequence 55, Application US/09944907
;; Publication No. US20020198147A1
;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin
;; APPLICANT: Batstein, David
;; APPLICANT: Eaton, Dan
;; APPLICANT: Ferrara, Napoleone
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gerritsen, Mary
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul
;; APPLICANT: Grimaldi, Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Hillan, Kenneth
;; APPLICANT: Kijavlin, Ivar
;; APPLICANT: Napier, Mary
;; APPLICANT: Roy, Margaret
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Wood, William
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; TITLE OF INVENTION: ACIDS ENCODING THE SAME
;; FILE REFERENCE: P2548P1C1
;; CURRENT APPLICATION NUMBER: US/09/944,907
;; CURRENT FILING DATE: 2001-08-31
;; PRIOR APPLICATION NUMBER: 09/866,028
;; PRIOR FILING DATE: 2001-05-25
;; NUMBER OF SEQ ID NOS: 120
;; SEQ ID NO 55
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-944-907-55

Query Match 71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. NO. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
||||:|
Db 658 QPGPVLKRP 666

RESULT 9
US-09-944-929-55

;; Sequence 55, Application US/09944929
;; Publication No. US20020197612A1
;; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin
;; APPLICANT: Batstein, David
;; APPLICANT: Eaton, Dan
;; APPLICANT: Ferrara, Napoleone
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gerritsen, Mary
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul
;; APPLICANT: Grimaldi, Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Hillan, Kenneth
;; APPLICANT: Kijavlin, Ivar
;; APPLICANT: Napier, Mary
;; APPLICANT: Roy, Margaret

```

; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944, 929
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 55
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-929-55

Query Match
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666

RESULT 10
US-10-174-590-60
; Sequence 60, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174, 590
; PRIOR FILING DATE: 2002-06-18
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-60

Query Match
Best Local Similarity 71.7%; Score 38; DB 9; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666

RESULT 11
US-10-176-758-60
; Sequence 60, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Godowski, Paul J.

```

```

; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176, 758
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-60

Query Match
Best Local Similarity 71.7%; Score 38; DB 9; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666

RESULT 12
US-10-175-737-60
; Sequence 60, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175, 737
; PRIOR FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-60

Query Match
Best Local Similarity 71.7%; Score 38; DB 9; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666

RESULT 13
US-10-173-706-60
; Sequence 60, Application US/10173706
; Publication No. US2003002293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian

```

```
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-60
```

```
Query Match          71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666
```

```
RESULT 14
US-10-175-738-60
; Sequence 60, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-60
```

```
Query Match          71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666
```

```
RESULT 15
US-10-175-752-60
; Sequence 60, Application US/10175752
; Publication No. US20030022295A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 60
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-60
```

```
Query Match          71.7%; Score 38; DB 9; Length 694;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 QPGPVLSP 10
Db 658 QPGPLKKP 666
```

```
Search completed: February 16, 2003, 22:02:34
Job time : 2.197 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 3.52891 Seconds
(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 AOPGPVLSQP 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCRNUS_COMB.pep:*

2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*

3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*

4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*

5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*

6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*

7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*

8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*

9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*

10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*

11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*

12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*

13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*

14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*

15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*

16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*

17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*

18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*

19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*

20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*

21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*

22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*

23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*

24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*

25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*

26: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*

27: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	685	27	US-60-243-468-1271
2	53	100.0	801	1	PCT-US01-08631-39827
3	53	100.0	843	1	PCT-US02-29964-410
4	53	100.0	852	21	US-09-763-909-2
5	53	100.0	865	1	PCT-US02-25829-25
6	43	81.1	71	1	PCT-US01-02723-214

7	43	81.1	71	21	US-09-707-351-214	Sequence 214, App
8	43	81.1	71	27	US-60-365-384-578	Sequence 578, App
9	42	79.2	348	1	PCT-US01-08631-44366	Sequence 254, App
10	42	79.2	348	1	PCT-US02-30474-2979	Sequence 44366, App
11	42	79.2	348	1	PCT-US02-30474-2979	Sequence 2979, App
12	42	79.2	348	27	US-60-324-631-2885	Sequence 2885, App
13	41	77.4	235	26	US-10-219-999-44244	Sequence 44244, A
14	41	77.4	250	21	US-09-708-427-54708	Sequence 54708, A
15	41	77.4	255	26	US-10-219-999-49842	Sequence 49842, A
16	41	77.4	310	26	US-10-219-999-38659	Sequence 38659, A
17	41	77.4	310	27	US-60-324-109-25624	Sequence 25624, A
18	41	77.4	316	21	US-09-708-427-54707	Sequence 54707, A
19	41	77.4	351	18	US-09-478-081-504	Sequence 504, App
20	41	77.4	353	19	US-09-513-996A-35642	Sequence 35642, A
21	41	77.4	353	21	US-09-708-427-1717	Sequence 1717, App
22	41	77.4	354	21	US-09-791-537-114482	Sequence 114482, A
23	41	77.4	357	26	US-10-219-999-51795	Sequence 51795, A
24	41	77.4	362	20	US-09-555-073-14	Sequence 14, App1
25	41	77.4	362	21	US-09-708-427-54706	Sequence 54706, A
26	41	77.4	362	26	US-10-219-999-33627	Sequence 33627, A
27	41	77.4	362	27	US-60-324-109-18893	Sequence 18893, A
28	41	77.4	365	19	US-09-513-996A-35641	Sequence 35641, A
29	41	77.4	365	19	US-09-573-655A-1205	Sequence 1205, App
30	41	77.4	365	19	US-09-573-655A-1205	Sequence 1205, App
31	41	77.4	365	21	US-09-708-427-1716	Sequence 1716, App
32	41	77.4	365	21	US-09-791-537-61865	Sequence 61865, App1
33	41	77.4	370	20	US-09-602-472-60	Sequence 127939, A
34	41	77.4	370	21	US-09-791-537-127939	Sequence 35640, A
35	41	77.4	397	19	US-09-513-996A-35640	Sequence 1715, App
36	41	77.4	397	21	US-09-708-427-1715	Sequence 1715, App
37	41	77.4	402	1	PCT-US02-27143-71	Sequence 71, App1
38	41	77.4	439	26	US-10-219-999-44281	Sequence 44281, A
39	41	77.4	439	27	US-60-324-109-24623	Sequence 24623, A
40	41	77.4	592	19	US-09-543-091-2	Sequence 2, App11
41	40	75.5	113	1	PCT-US01-01338-1677	Sequence 1677, App
42	40	75.5	113	21	US-09-764-877-1677	Sequence 1677, App
43	40	75.5	113	26	US-10-242-515-1677	Sequence 1677, App
44	40	75.5	199	1	PCT-US01-08631-50835	Sequence 50835, A
45	40	75.5	247	18	US-09-417-507-28769	Sequence 28769, A

ALIGNMENTS

RESULT 1

US-60-243-468-1271

Sequence 1271, Application US/60243468

GENERAL INFORMATION:

APPLICANT: Beasley, Ellen

TITLE OF INVENTION: ISOLATED HUMAN DRUG TARGET PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG TARGET PROTEINS,

TITLE OF INVENTION: AND USES THEREOF

FILE REFERENCE: C1000929

CURRENT APPLICATION NUMBER: US/60/243,468

CURRENT FILING DATE: 2000-10-27

NUMBER OF SEQ ID NOS: 2121

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1271

LENGTH: 685

TYPE: PRT

ORGANISM: HUMAN

US-60-243-468-1271

Query Match 100.0%; Score 53; DB 27; Length 685;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSQP 10

DB 473 AOPGPVLSQP 482

RESULT 2

PCT-US01-08631-39827
; Sequence 39827, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO: 39827
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-39827

Query Match 100.0%; Score 53; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
DB 443 AOPGPVLSOP 452

RESULT 3
PCT-US02-29964-410
; Sequence 410, Application PC/TUS0229964
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qiang A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aiding J.
; APPLICANT: Wang, Dunrui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Asundi, Vinod
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Weng, Gezhi
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 809ACIP PCT
; CURRENT APPLICATION NUMBER: PCT/US02/29964
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 60/323,739
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28

Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 992
; SOFTWARE: PC_FL_genes Version 6.0
; SEQ ID NO: 410
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-29964-410

Query Match 100.0%; Score 53; DB 1; Length 843;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
DB 485 AOPGPVLSOP 494

RESULT 4
US-09-763-909-2
; Sequence 2, Application US/09763909
; GENERAL INFORMATION:
; APPLICANT: Dikstein, Rivka
; APPLICANT: Yamle-Hezi, Avaya
; TITLE OF INVENTION: A TRANSCRIPTION FACTOR TFIIID SUBUNIT,
; TITLE OF INVENTION: TAFII105, POLYPEPTIDES, DNA ENCODING THEREFOR AND
; FILE REFERENCE: 13005/002001
; CURRENT APPLICATION NUMBER: US/09/763,909
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-763-909-2

Query Match 100.0%; Score 53; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
DB 443 AOPGPVLSOP 452

RESULT 5
PCT-US02-25829-25
; Sequence 25, Application PC/TUS0225829
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: BAROSSO, Ines
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: BLAKE, Julie J.
; APPLICANT: BOROFSKY, Mark L.
; APPLICANT: BUREFORD, Neil
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: FORSTHE, Ian J.
; APPLICANT: GIETZEN, Kimberly J.
; APPLICANT: GORVAD, Ann E.
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: HAPALIA, April J.A.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: ISON, Craig H.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: LAL, Preeti G.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: LEE, Sally
; APPLICANT: LEE, Soo Yeun

APPLICANT: LI, Joana X.
APPLICANT: LU, Yung Aina M.
APPLICANT: LU, Yan
APPLICANT: LEHR-MASON, Patricia M.
APPLICANT: NGUYEN, Daniel B.
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: SPRAGUE, William W.
APPLICANT: TANG, Y. Tom
APPLICANT: THANGAVELU, Kavitha
APPLICANT: THORNTON, Michael
APPLICANT: TRAN, Uyen K.
APPLICANT: WALIA, Narinder K.
APPLICANT: WARREN, Bridget A.
APPLICANT: XU, Yuming
APPLICANT: YAO, Monique G.
APPLICANT: YUE, Henry
APPLICANT: YUE, Huibin
APPLICANT: ZEBARADIAN, Yeganeh
TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
FILE REFERENCE: PF-1146 PCT
CURRENT APPLICATION NUMBER: PCT/US02/25829
CURRENT FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: US 60/313,111
PRIOR FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 60/314,662
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/314,756
PRIOR FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/315,105
PRIOR FILING DATE: 2001-08-27
PRIOR APPLICATION NUMBER: US 60/316,751
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/316,856
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,185
PRIOR FILING DATE: 2001-10-05
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PERL Program
SEQ ID NO 25
LENGTH: 865
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: incyte ID No: 4398735CD1
PCT-US02-25829-25
Query Match 100.0%; Score 53; DB 1; Length 865;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLSOP 10
Db 512 AOPGPVLSOP 521
RESULT 6
PCT-US01-02723-214
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
APPLICANT: Yeung, George
APPLICANT: Ford, John E.
APPLICANT: Boyle, Bryan J.
APPLICANT: Arterburn, Matthew C.
APPLICANT: Dirmanc, Radoje T.
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhou, Ping
APPLICANT: Wehrman, Thomas
TITLE OF INVENTION: NOVEL FETAL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-020

CURRENT APPLICATION NUMBER: PCT/US01/02723
CURRENT FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 09/707,351
PRIOR FILING DATE: 2000-11-06
PRIOR APPLICATION NUMBER: 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/663,870
PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 1262
SOFTWARE: PatentIn version 3.0
SEQ ID NO 214
LENGTH: 71
TYPE: PRT
ORGANISM: homo sapiens
PCT-US01-02723-214
Query Match 81.1%; Score 43; DB 1; Length 71;
Best Local Similarity 70.0%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLSOP 10
Db 40 SRPGVLSOP 49
RESULT 7
US-09-707-351-214
GENERAL INFORMATION:
APPLICANT: Yeung, George
APPLICANT: Ford, John E.
APPLICANT: Boyle, Bryan J.
APPLICANT: Arterburn, Matthew C.
APPLICANT: Dirmanc, Radoje T.
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhou, Ping
APPLICANT: Wehrman, Thomas
TITLE OF INVENTION: NOVEL FETAL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 796
CURRENT APPLICATION NUMBER: US/09/707,351
CURRENT FILING DATE: 2000-11-06
PRIOR APPLICATION NUMBER: 09/663,870
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/491,404
PRIOR FILING DATE: 2000-01-25
NUMBER OF SEQ ID NOS: 954
SOFTWARE: PatentIn version 3.0
SEQ ID NO 214
LENGTH: 71
TYPE: PRT
ORGANISM: homo sapiens
US-09-707-351-214
Query Match 81.1%; Score 43; DB 21; Length 71;
Best Local Similarity 70.0%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLSOP 10
Db 40 SRPGVLSOP 49
RESULT 8
US-60-365-384-578
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Jilan-Tui
APPLICANT: Wang, Zhiwei
APPLICANT: Zhang, Jie
APPLICANT: Zhou, Ping

APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Ghosh, Malabika
APPLICANT: Zhao, Qing A.
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aidong J.
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Weng, Gezhi
TITLE OF INVENTION: Novel Nucleic Acids and
POLYPEPTIDES
FILE REFERENCE: 814
CURRENT APPLICATION NUMBER: US/60/365,384
CURRENT FILING DATE: 2002-03-14
NUMBER OF SEQ ID NOS: 666
SOFTWARE: PC_FL_genes Version 6.0
SEQ ID NO 578
LENGTH: 71
TYPE: PRT
ORGANISM: Homo sapiens
US-60-365-384-578

Query Match 81.1%; Score 43; DB 27; Length 71;
Best Local Similarity 70.0%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AOPGPVLSOP 10
Db 40 SRGPVLSER 49

RESULT 9
US-60-365-384-254
Sequence 254, Application US/60365384
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Wang, Zhiwei
APPLICANT: Zhang, Jie
APPLICANT: Zhou, Ping
APPLICANT: Wehrman, Tom
APPLICANT: Wang, Jian-Rui
APPLICANT: Ghosh, Malabika
APPLICANT: Zhao, Qing A.
APPLICANT: Asundi, Vinod
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aidong J.
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Weng, Gezhi
TITLE OF INVENTION: Novel Nucleic Acids and
POLYPEPTIDES
FILE REFERENCE: 814
CURRENT APPLICATION NUMBER: US/60/365,384
CURRENT FILING DATE: 2002-03-14
NUMBER OF SEQ ID NOS: 666
SOFTWARE: PC_FL_genes Version 6.0
SEQ ID NO 254
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
US-60-365-384-254

Query Match 81.1%; Score 43; DB 27; Length 91;
Best Local Similarity 70.0%; Pred. No. 54;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AOPGPVLSOP 10
Db 40 SRGPVLSER 49

RESULT 10
PCT-US01-08631-44366
Sequence 44366, Application PC/TUS0108631
GENERAL INFORMATION:
APPLICANT: Hysq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-049
CURRENT APPLICATION NUMBER: PCT/US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO 44366
LENGTH: 348
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(348)
OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08631-44366

Query Match 79.2%; Score 42; DB 1; Length 348;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AOPGPVLSOP 10
Db 99 AOPGPVLSOP 108

RESULT 11
PCT-US02-30474-2979
Sequence 2979, Application PC/TUS0230474
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Xue, Aidong J.
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Dunrui
APPLICANT: Ma, Yungqing
APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhiwei
APPLICANT: Weng, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Drmanac, Radolje T
TITLE OF INVENTION: Novel Nucleic Acids and
POLYPEPTIDES
FILE REFERENCE: 810CIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/30474
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: US 60/324,631
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03

;; PRIOR APPLICATION NUMBER: US 09/560,875
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: PCT/US01/03800
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 09/515,126
;; PRIOR FILING DATE: 2000-02-28
;; Remaining prior application data removed - See file Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 3476
;; SOFTWARE: PL_FL_genes Version 6.0
;; SEQ ID NO 2979
;; LENGTH: 348
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc.feature
;; LOCATION: (1)...(348)
;; OTHER INFORMATION: Xaa - any amino acid or symbol as shown in the table 8 as set for
PCT-US02-30474-2379

Query Match 79.2%; Score 42; DB 1; Length 348;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
Db 99 AOPGPVLSOP 108

RESULT 12
US-60-324-631-2885

;; Sequence 2885, Application US/60324631

;; GENERAL INFORMATION:

;; APPLICANT: Tang, Y. Tom

;; APPLICANT: Zhang, Jie

;; APPLICANT: Ren, Feiyan

;; APPLICANT: Xue, Aidong J.

;; APPLICANT: Zhao, Qing A.

;; APPLICANT: Wang, Jian-Rui

;; APPLICANT: Mehrman, Tom

;; APPLICANT: Zhou, Ping

;; APPLICANT: Ghosh, Malabika

;; APPLICANT: Wang, Dunrui

;; APPLICANT: Ma, Yungling

;; APPLICANT: Asundi, Vinod

;; APPLICANT: Wang, Zhilwei

;; APPLICANT: Meng, Gezhi

;; APPLICANT: Haley-Vicente, Dana

;; TITLE OF INVENTION: Novel Nucleic Acids and

;; FILE REFERENCE: 810

;; CURRENT APPLICATION NUMBER: US/60/324,631

;; PRIOR APPLICATION NUMBER: 2001-09-24

;; PRIOR FILING DATE: 2000-01-21

;; PRIOR APPLICATION NUMBER: US 09/552,317

;; PRIOR FILING DATE: 2000-04-25

;; PRIOR APPLICATION NUMBER: PCT/US00/35017

;; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: US 09/491,404

;; PRIOR FILING DATE: 2000-01-25

;; PRIOR APPLICATION NUMBER: PCT/US01/02623

;; PRIOR FILING DATE: 2001-01-25

;; PRIOR APPLICATION NUMBER: US 09/496,914

;; PRIOR FILING DATE: 2000-02-03

;; PRIOR APPLICATION NUMBER: US 09/560,875

;; PRIOR FILING DATE: 2000-04-27

;; PRIOR APPLICATION NUMBER: PCT/US01/03800

;; PRIOR FILING DATE: 2001-02-05

;; PRIOR APPLICATION NUMBER: US 09/515,126

;; PRIOR FILING DATE: 2000-02-28

;; PRIOR APPLICATION NUMBER: US 09/577,409

;; PRIOR FILING DATE: 2000-05-18

;; PRIOR APPLICATION NUMBER: PCT/US01/04927
;; PRIOR FILING DATE: 2001-02-26
;; PRIOR APPLICATION NUMBER: US 09/519,705
;; PRIOR FILING DATE: 2000-03-07
;; PRIOR APPLICATION NUMBER: US 09/574,454
;; PRIOR FILING DATE: 2000-05-19
;; PRIOR APPLICATION NUMBER: PCT/US01/04941
;; PRIOR FILING DATE: 2001-03-05
;; PRIOR APPLICATION NUMBER: US 09/540,217
;; PRIOR FILING DATE: 2000-03-31
;; PRIOR APPLICATION NUMBER: US 09/649,167
;; PRIOR FILING DATE: 2000-08-23
;; PRIOR APPLICATION NUMBER: PCT/US01/08631
;; PRIOR FILING DATE: 2001-03-30
;; PRIOR APPLICATION NUMBER: US 09/552,929
;; PRIOR FILING DATE: 2000-04-18
;; PRIOR APPLICATION NUMBER: US 09/770,160
;; PRIOR FILING DATE: 2001-01-26
;; PRIOR APPLICATION NUMBER: PCT/US01/08656
;; PRIOR FILING DATE: 2001-04-18
;; PRIOR APPLICATION NUMBER: US 09/577,408
;; PRIOR FILING DATE: 2000-05-18
;; PRIOR APPLICATION NUMBER: PCT/US01/14827
;; PRIOR FILING DATE: 2001-05-16
;; NUMBER OF SEQ ID NOS: 3334
;; SOFTWARE: PL_FL_genes Version 6.0
;; SEQ ID NO 2885
;; LENGTH: 348
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc.feature
;; LOCATION: (1)...(348)
;; OTHER INFORMATION: Xaa - any amino acid or symbol as shown in the table 8 as set
US-60-324-631-2885

Query Match 79.2%; Score 42; DB 27; Length 348;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AOPGPVLSOP 10
Db 99 AOPGPVLSOP 108

RESULT 13
US-10-219-999-44244

;; Sequence 44244, Application US/10219999

;; GENERAL INFORMATION:

;; APPLICANT: Cao, Yongwei

;; APPLICANT: Edgerton, Michael D

;; APPLICANT: Hinkle, Gregory J.

;; APPLICANT: Kovalic, David K.

;; APPLICANT: Liu, Jingdong

;; APPLICANT: Stein, Joshua

;; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT

;; FILE REFERENCE: 38-10(52726)C

;; CURRENT APPLICATION NUMBER: US/10/219,999

;; PRIOR APPLICATION NUMBER: 2002-08-15

;; PRIOR FILING DATE: 2000-03-24, 109

;; PRIOR APPLICATION NUMBER: US 60/324,109

;; PRIOR FILING DATE: 2001-09-21

;; PRIOR APPLICATION NUMBER: US 60/312,544

;; PRIOR FILING DATE: 2001-08-15

;; NUMBER OF SEQ ID NOS: 63520

;; SEQ ID NO 44244

;; LENGTH: 235

;; TYPE: PRT

;; ORGANISM: Zea mays

;; US-10-219-999-44244

Query Match 77.4%; Score 41; DB 26; Length 235;

Best Local Similarity 100.0%; Pred. No. 3.2e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLS 8
Db 23 AOPGPVLS 30

Search completed: February 16, 2003, 22:18:32
Job time : 4.52891 secs

RESULT 14

US-09-708-427-54708
; Sequence 54708, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54708
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..250
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..250
; OTHER INFORMATION: Ceres Seq. ID 1933788
US-09-708-427-54708

Query Match

Best Local Similarity 77.4%; Score 41; DB 21; Length 250;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLS 8
Db 48 AOPGPVLS 55

RESULT 15

US-10-219-999-49842
; Sequence 49842, Application US/10219999
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Liu, Jindong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)C
; CURRENT APPLICATION NUMBER: US/10/219,999
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/324,109
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/312,544
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 63520
; SEQ ID NO 49842
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Zea mays
US-10-219-999-49842

Query Match

Best Local Similarity 77.4%; Score 41; DB 26; Length 255;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AOPGPVLS 8
Db 56 AOPGPVLS 63

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:00:19 ; Search time 0.406852 seconds
(without alignments)
2076.603 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 AOPGPVLSQP 10

Scoring table: BL0SUM62

Searched: Gap0 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 429898

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_US02-32727-6510
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	83.0	136	1 PCT-US02-32727-6510	Sequence 6510, Ap
2	44	83.0	136	US-10-057-498-6510	Sequence 6510, Ap
3	41	77.4	370	US-09-602-472A-60	Sequence 60, Appl
4	39	73.6	775	1 PCT-US02-10366-58	Sequence 58, Appl
5	38	71.7	608	US-10-351-334-213	Sequence 213, App
6	38	71.7	694	US-10-125-923A-60	Sequence 60, Appl
7	38	71.7	694	US-10-205-892-60	Sequence 60, Appl
8	38	71.7	694	US-10-174-575-60	Sequence 60, Appl
9	38	71.7	694	US-10-174-575A-60	Sequence 60, Appl
10	38	71.7	694	US-10-187-755-60	Sequence 60, Appl
11	38	71.7	694	US-10-187-749-60	Sequence 60, Appl
12	38	71.7	694	US-10-199-672-60	Sequence 60, Appl
13	38	71.7	694	US-10-194-486-60	Sequence 60, Appl
14	38	71.7	716	US-10-351-334-370	Sequence 370, App
15	37	69.8	194	US-10-299-003-52	Sequence 52, Appl
16	37	69.8	222	US-09-724-676-69050	Sequence 69050, A
17	37	69.8	222	US-09-724-676-69051	Sequence 69050, A
18	37	69.8	222	US-09-724-676A-69050	Sequence 69050, A
19	37	69.8	222	US-09-724-676A-69051	Sequence 69051, A
20	37	69.8	234	US-09-724-676-69052	Sequence 69052, A
21	37	69.8	234	US-09-724-676A-69052	Sequence 69052, A
22	37	69.8	333	US-09-724-676-64147	Sequence 64147, A
23	37	69.8	333	US-09-724-676A-64147	Sequence 64147, A
24	37	69.8	559	US-10-289-003-15	Sequence 15, Appl
25	37	69.8	756	US-10-338-462-9	Sequence 9, Appl
26	37	69.8	768	US-09-724-676-69047	Sequence 69047, A

27	37	69.8	768	5 US-09-724-676A-69047	Sequence 69047, A
28	37	69.8	967	6 US-10-017-161-1766	Sequence 1766, Ap
29	37	69.8	972	6 US-10-276-781-1513	Sequence 1513, Ap
30	37	69.8	972	6 US-10-276-781-1535	Sequence 1535, Ap
31	36	67.9	183	6 US-10-264-237-2472	Sequence 2472, Ap
32	36	67.9	190	6 US-10-203-138A-13166	Sequence 13166, A
33	36	67.9	362	5 US-09-373-902-7	Sequence 7, Appl
34	36	67.9	422	6 US-10-264-237-2090	Sequence 2090, Ap
35	36	67.9	1109	1 PCT-US02-40059-8	Sequence 8, Appl
36	36	67.9	1220	6 US-10-144-779-525	Sequence 525, App
37	36	67.9	2507	1 PCT-US02-40059-2	Sequence 2, Appl
38	35	66.0	53	6 US-10-203-138A-11392	Sequence 11392, A
39	35	66.0	72	1 PCT-US02-32821-23821	Sequence 23821, A
40	35	66.0	72	6 US-10-057-498-23821	Sequence 23821, A
41	35	66.0	172	1 PCT-US02-32850-5	Sequence 5, Appl
42	35	66.0	172	1 PCT-US02-32850A-5	Sequence 5, Appl
43	35	66.0	350	1 PCT-US02-32874-3	Sequence 3, Appl
44	35	66.0	350	5 US-09-906-777B-236	Sequence 236, App
45	35	66.0	350	5 US-09-904-011C-236	Sequence 236, App

ALIGNMENTS

```
RESULT 1
PCT-US02-32727-6510
Sequence 6510, Application PC/TUS0232727
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Persing, David
APPLICANT: Bhatia, Ajay
APPLICANT: Malsomeuve, Jean Francois
APPLICANT: Zhang, Yanni
APPLICANT: Wang, Siqing
APPLICANT: Jen, Shiyian
APPLICANT: Lodes, Michael
APPLICANT: Benson, Darin
APPLICANT: Jones, Robert
APPLICANT: Carter, Darick
APPLICANT: Barth, Brenda
APPLICANT: Douglas, John
TITLE OF INVENTION: Compositions and Methods for the therapy and Diagnosis of Acne
FILE REFERENCE: 210121.514C1
CURRENT APPLICATION NUMBER: PCT/US02/32727
CURRENT FILING DATE: 2002-10-11
NUMBER OF SEQ ID NOS: 30992
SEQ ID NO 6510
LENGTH: 136
TYPE: PRT
ORGANISM: Propionibacterium
PCT-US02-32727-6510
Query Match 83.0%; Score 44; DB 1; Length 136;
Best local Similarity 80.0%; Pred. No. 6.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 61 AOPGPVLSQP 10
61 AOPGPVLSQP 70
RESULT 2
US-10-057-498-6510
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer
APPLICANT: Skeiky, Yasir
APPLICANT: Persing, David
TITLE OF INVENTION: Compositions and Methods for the therapy and Diagnosis of Acne
FILE REFERENCE: 210121.514
CURRENT APPLICATION NUMBER: US/10/057,498
CURRENT FILING DATE: 2001-04-20
```

NUMBER OF SEQ ID NOS: 29212
SEQ ID NO 6510
LENGTH: 136
TYPE: PRT
ORGANISM: Propionibacterium
US-10-057-498-6510

Query Match
Best Local Similarity 83.0%; Score 44; DB 6; Length 136;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 ACPGVLSQ 10
Db 61 ACPGVLSQ 70

RESULT 3
US-09-602-472A-60
Sequence 60, Application US/09602472A
GENERAL INFORMATION:
APPLICANT: Bidney, Dennis L.
APPLICANT: Hu, Xu
APPLICANT: Li, Guohua
TITLE OF INVENTION: Sunflower Disease Resistance Genes
FILE REFERENCE: 35718/200630
CURRENT APPLICATION NUMBER: US/09/602,472A
CURRENT FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 60/140,876
PRIOR FILING DATE: 1999-06-24
NUMBER OF SEQ ID NOS: 64
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 60
LENGTH: 370
TYPE: PRT
ORGANISM: Lycopersicon esculentum
US-09-602-472A-60

Query Match
Best Local Similarity 77.4%; Score 41; DB 5; Length 370;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACPGVLSQ 8
Db 161 ACPGVLSQ 168

RESULT 4
PCT-US02-10366-58
Sequence 58, Application PC/TUS0210366
GENERAL INFORMATION:
APPLICANT: Padigaru, Murallidhara
APPLICANT: Shenoy, Suresh
APPLICANT: Kekuda, Ramesh
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter
APPLICANT: Smithson, Glendina
APPLICANT: Guo, Xiaojia
APPLICANT: Gerlach, Valerie
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Ferenc
APPLICANT: Li, Li
APPLICANT: Zernusen, Bryan
APPLICANT: Tchernev, Vellizar
APPLICANT: Gangoli, Eshe
APPLICANT: Vernet, Corine
APPLICANT: Spytek, Kimberly
APPLICANT: Malvankar, Uriel
APPLICANT: Patursajan, Meera
APPLICANT: Miller, Charles
APPLICANT: Taupier, Raymond J. Jr.
APPLICANT: Heyes, Melvyn
APPLICANT: Ju, Jiongfang
APPLICANT: Peyman, John

APPLICANT: Catterton, Elina
APPLICANT: MacDougall, John
APPLICANT: Edinger, Shlomit
APPLICANT: Stone, David
APPLICANT: Mazur, Ann
TITLE OF INVENTION: NOVEL ANTIBODIES THAT BIND TO ANTIGENIC POLYPEPTIDES,
TITLE OF INVENTION: ACIDS ENCODING THE ANTIGENS, AND METHODS OF USE
FILE REFERENCE: 21402-322A-061
CURRENT APPLICATION NUMBER: PCT/US02/10366
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 10/114153
PRIOR FILING DATE: 2002-04-02
PRIOR APPLICATION NUMBER: 60/281086
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/281906
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/282020
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: 60/282930
PRIOR FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: 60/283512
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283444
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/283657
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283710
PRIOR FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/283678
PRIOR FILING DATE: 2001-04-13
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 251
SEQ ID NO 58
LENGTH: 775
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-10366-58

Query Match
Best Local Similarity 73.6%; Score 39; DB 1; Length 775;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ACPGVLSQ 9
Db 264 ACPGVLSQ 272

RESULT 5
US-10-351-334-213
Sequence 213, Application US/10351334
GENERAL INFORMATION:
APPLICANT: Komatsoulis et al
TITLE OF INVENTION: 98 Human Secreted Proteins
FILE REFERENCE: P2031P2
CURRENT APPLICATION NUMBER: US/10/351,334
CURRENT FILING DATE: 2003-01-27
PRIOR APPLICATION NUMBER: 60/350,898
PRIOR FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: 09/489,847
PRIOR FILING DATE: 2000-01-24
PRIOR APPLICATION NUMBER: PCT/US99/17130
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: 60/094,657
PRIOR FILING DATE: 1998-07-30
PRIOR APPLICATION NUMBER: 60/095,486
PRIOR FILING DATE: 1998-08-05
PRIOR APPLICATION NUMBER: 60/096,319
PRIOR FILING DATE: 1998-08-12
PRIOR APPLICATION NUMBER: 60/095,454
PRIOR FILING DATE: 1998-08-06
PRIOR APPLICATION NUMBER: 60/095,455
PRIOR FILING DATE: 1998-08-06

NUMBER OF SEQ ID NOS: 376
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 213
LENGTH: 608
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (265)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (597)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-351-334-213

Query Match 71.7%; Score 38; DB 6; Length 608;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGVLSOP 10
||||:|:|
Db 353 QPGPLKKP 361

RESULT 6
US-10-125-923A-60
Sequence 60, Application US/10125923A
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jia
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C79
CURRENT APPLICATION NUMBER: US/10/125,923A
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 60
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Sapien
US-10-125-923A-60

Query Match 71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGVLSOP 10
||||:|:|
Db 658 QPGPLKKP 666

RESULT 7
US-10-205-892-60
Sequence 60, Application US/10205892
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jia
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C517
CURRENT APPLICATION NUMBER: US/10/205,892
PRIOR FILING DATE: 2002-07-26
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 60
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Sapien
US-10-205-892-60

Query Match 71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 QPGVLSOP 10
||||:|:|
Db 658 QPGPLKKP 666

RESULT 8
US-10-174-575-60
Sequence 60, Application US/10174575
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jia
APPLICANT: Desnoyers, Luc

;; PRIOR FILING DATE: 1997-10-28
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 60
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-187-755-60

Query Match 71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
||||:|:
DB 658 QPGPLKKP 666

RESULT 11
US-10-187-749-60
;; Sequence 60, Application US/10187749
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Chen, Jian
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Smith, Victoria
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Wood, William I.
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3430R1C1
;; CURRENT APPLICATION NUMBER: US/10/187,749
;; CURRENT FILING DATE: 2002-07-01
;; PRIOR APPLICATION NUMBER: US/10/052,586
;; PRIOR FILING DATE: 2002-01-15
;; PRIOR APPLICATION NUMBER: 60/059263
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/059266
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/063120
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063121
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063486
;; PRIOR FILING DATE: 1997-10-21
;; PRIOR APPLICATION NUMBER: 60/063540
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063541
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063544
;; PRIOR FILING DATE: 1997-10-28
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 60
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-187-749-60

Query Match 71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
||||:|:
DB 658 QPGPLKKP 666

RESULT 12
US-10-199-672-60
;; Sequence 60, Application US/10199672
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Chen, Jian
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Smith, Victoria
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Wood, William I.
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3430R1C1
;; CURRENT APPLICATION NUMBER: US/10/199,672
;; CURRENT FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: US/10/052,586
;; PRIOR FILING DATE: 2002-01-15
;; PRIOR APPLICATION NUMBER: 60/059263
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/059266
;; PRIOR FILING DATE: 1997-09-18
;; PRIOR APPLICATION NUMBER: 60/062250
;; PRIOR FILING DATE: 1997-10-17
;; PRIOR APPLICATION NUMBER: 60/063120
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063121
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/063486
;; PRIOR FILING DATE: 1997-10-21
;; PRIOR APPLICATION NUMBER: 60/063540
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063541
;; PRIOR FILING DATE: 1997-10-28
;; PRIOR APPLICATION NUMBER: 60/063544
;; PRIOR FILING DATE: 1997-10-28
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 612
;; SEQ ID NO 60
;; LENGTH: 694
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-199-672-60

Query Match 71.7%; Score 38; DB 6; Length 694;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSP 10
||||:|:
DB 658 QPGPLKKP 666

RESULT 13
US-10-194-486-60
;; Sequence 60, Application US/10194486
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Chen, Jian
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Pan, James
;; APPLICANT: Smith, Victoria
;; APPLICANT: Watanabe, Colin K.
;; APPLICANT: Wood, William I.

APPLICANT: Zhang Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/194,486
CURRENT FILING DATE: 2002-07-11
PRIOR APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
SEQ ID NO 60
LENGTH: 694
TYPE: PRT
ORGANISM: Homo Sapien
US-10-194-486-60

Query Match
Best Local Similarity 71.7%; Score 38; DB 6; Length 694;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGPEVLSQP 10
DB 658 QGPEVLSKP 666

RESULT 14
US-10-351-334-370
Sequence 370, Application US/10351334
GENERAL INFORMATION:
APPLICANT: Komatsoulis et al
TITLE OF INVENTION: 98 Human Secreted Proteins
FILE REFERENCE: P2031P2
CURRENT APPLICATION NUMBER: US/10/351,334
CURRENT FILING DATE: 2003-01-27
PRIOR APPLICATION NUMBER: 60/350,898
PRIOR FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: 09/489,847
PRIOR FILING DATE: 2000-01-24
PRIOR APPLICATION NUMBER: PCT/US99/17130
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: 60/094,657
PRIOR FILING DATE: 1998-07-30
PRIOR APPLICATION NUMBER: 60/095,486
PRIOR FILING DATE: 1998-08-05
PRIOR APPLICATION NUMBER: 60/096,319
PRIOR FILING DATE: 1998-08-12
PRIOR APPLICATION NUMBER: 60/095,454
PRIOR FILING DATE: 1998-08-06
PRIOR APPLICATION NUMBER: 60/095,455
PRIOR FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 376
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 370
LENGTH: 716
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (373)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (705)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-351-334-370

Query Match
Best Local Similarity 71.7%; Score 38; DB 6; Length 716;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGPEVLSQP 10
DB 461 QGPEVLSKP 469

RESULT 15
US-10-299-003-52
Sequence 52, Application US/10299003
GENERAL INFORMATION:
APPLICANT: Steck, Peter
APPLICANT: Pershouse, Mark A.
APPLICANT: Jasser, Samar
APPLICANT: Tung, Alfred W.K.
APPLICANT: Tavligian, Sean V.
TITLE OF INVENTION: A TUMOR SUPPRESSOR DESIGNATED TS10Q23.3
FILE REFERENCE: 2318-385
CURRENT APPLICATION NUMBER: US/10/299,003
CURRENT FILING DATE: 2002-11-19
PRIOR APPLICATION NUMBER: U.S. 09/140,749
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: U.S. 08/791,115
PRIOR FILING DATE: 1997-01-30
PRIOR APPLICATION NUMBER: U.S. 60/057,750
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: U.S. 60/083,563
PRIOR FILING DATE: 1998-04-30
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 52
LENGTH: 194
TYPE: PRT
ORGANISM: Mus musculus
US-10-299-003-52

Query Match
Best Local Similarity 69.8%; Score 37; DB 6; Length 194;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 QGPEVLSQP 10
DB 78 QGPEVLSKP 85

Search completed: February 16, 2003, 22:04:26
Job time : 1.40685 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 0.33833 Seconds
(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53

Sequence: 1 AQPGLVLSQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	77.4	365	2 T02181	probable protein kinase
2	38	71.7	987	2 A75496	conserved hypothe
3	37	69.8	349	2 D83048	hypothetical prote
4	37	69.8	891	2 T22560	hypothetical prote
5	36	67.9	163	2 B82708	hypothetical prote
6	36	67.9	233	2 H72648	hypothetical prote
7	36	67.9	308	1 QOLJX1	trans-activating t
8	36	67.9	361	2 S59531	manganese peroxid
9	36	67.9	603	2 A75373	probable N-acetyl
10	36	67.9	686	2 F82871	probable FMN oxido
11	35.5	67.0	1071	2 B84062	hypothetical prote
12	35	66.0	188	2 B45540	coat protein - phy
13	35	66.0	190	2 S41553	coat protein - bel
14	35	66.0	309	2 E36596	nuclear factor I f
15	35	66.0	311	2 B86839	probable isopenten
16	35	66.0	328	2 T17789	chitosanase homolo
17	35	66.0	339	2 D41377	methanol dehydroge
18	35	66.0	350	2 JC7188	REIC protein - hum
19	35	66.0	357	2 PC4293	nuclear factor 1 f
20	35	66.0	365	2 D84848	probable protein k
21	35	66.0	371	2 A10199	tRNA (5-methylamin
22	35	66.0	386	2 S43569	nuclear factor I-A
23	35	66.0	446	2 C69653	hydrolytic enzyme
24	35	66.0	461	2 C69653	nuclear factor I-A
25	35	66.0	498	2 S45567	nuclear factor I-A
26	35	66.0	509	2 S45565	nuclear factor I-A
27	35	66.0	509	2 JC5428	nuclear factor I f
28	35	66.0	522	2 S09996	nuclear factor I f
29	35	66.0	532	2 B36596	nuclear factor I -

30	35	66.0	587	2 JC5300	Ran GTPase activat
31	35	66.0	589	2 A36983	RNAI homolog fung1
32	35	66.0	589	2 T52070	RNAI protein homol
33	35	66.0	1075	2 G59434	KRAA0411 protein l
34	35	66.0	1108	2 T35827	probable membrane
35	35	66.0	1285	2 T14171	ataxin-2 - mouse
36	35	66.0	1445	2 T10728	Basoon protein -
37	35	66.0	3942	2 T42730	probable gag/pol p
38	34	64.2	124	2 T26799	hypothetical prote
39	34	64.2	135	2 C83092	stringent starvati
40	34	64.2	152	2 T35798	hypothetical prote
41	34	64.2	188	2 P00180	CyB protein - Syn
42	34	64.2	236	1 A30225	regulatory protein
43	34	64.2	236	2 T51073	regulatory protein
44	34	64.2	298	2 S27965	hypothetical prote
45	34	64.2	315	2 T49982	hypothetical prote

ALIGNMENTS

RESULT 1
T02181
probable protein kinase [imported] - Arabidopsis thaliana
N:Alternate names: protein kinase homolog F14M4.11
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02181; F84910
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, September 1998
A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.
A:Reference number: Z14609
A:Accession: T02181
A>Status: translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-365 <R0U>
A:Cross-references: EMBL:AC004411; NID:93522932; PID:93522961
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84910
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-365 <STO>
A:Cross-references: GB:AE002093; NID:93522961; PID:AMC34243.1; GSPDB:GN00139
C:Genetics:
A:Gene: F14M4.11; At2g47060
A:Map position: 2
A:Introns: 32/1; 116/3; 159/1; 250/2; 294/3; 323/3
C:Superfamily: Kinase-related transforming protein; protein kinase homology

Query Match 77.4%; Score 41; DB 2; Length 365;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQPGLVLS 8
Db 166 AQPGLVLS 173

RESULT 2
A75496
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: A75496
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Yamahavean, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: AF5250; MUID:20036896; PMID:10567266
A:Accession: AF5496
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-987 <MHI>
A:Cross-references: GB:AE001920; GB:AE000513; NID:96458318; PIDN:AAF10199.1; PID:9645832
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0620
A:Map position: 1

Query Match 71.7%; Score 38; DB 2; Length 987;
Best Local Similarity 70.0%; Pred. No. 80;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AOPGVLSP 10
| | | | | : |
Db 488 AOPGVLSP 497

RESULT 3
D83048
hypothetical protein PA4774 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83048
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lam,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10964043
A:Accession: D83048
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-349 <STO>
A:Cross-references: GB:AE004891; GB:AE004091; NID:99591037; PIDN:AA08160.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4774

Query Match 69.8%; Score 37; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AOPGVL 7
| | | | |
Db 315 AOPGVL 321

RESULT 4
T22560
hypothetical protein F53C11.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22560
R:Baynes, C.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z19581
A:Accession: T22560
A:Status: Preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-891 <WII>
A:Cross-references: EMBL:Z79756; PIDN:CA802122.1; GSPDB:GN00023; CESP:F53C11.5
A:Experimental source: clone F53C11
C:Genetics:
A:Gene: CESP:F53C11.5
A:Map position: 5
A:Introns: 25/3; 59/1; 137/1; 287/1; 313/1; 343/3; 420/1; 455/2; 637/2; 708/2; 831/2; 86
Query Match 69.8%; Score 37; DB 2; Length 891;

Best Local Similarity 66.7%; Pred. No. 11e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 2 OPGVLSOP 10
| | | | |
Db 49 OPGVLSOP 57

RESULT 5
B82708
hypothetical protein XP1226 [imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: B82708
R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: AB2515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82708
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-163 <SIM>
A:Cross-references: GB:AE003956; GB:AE003849; NID:99106196; PIDN:AA04036.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Stimpson, A.J.G.; Reimach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
B:Trons, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Doiry, H.; Facincani, A.P.; Ferreira, A.D.S.
submitted to GeneBank, June 2000
A:Authors: Ferreira, V.C.R.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurama, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeiri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1226

Query Match 67.9%; Score 36; DB 2; Length 163;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 OPGVLSOP 10
| | | | |
Db 58 OPGVLSOP 65

RESULT 6
H72648
hypothetical protein APE0622 - *Aeropyrum pernix* (strain K1)
C:Species: *Aeropyrum pernix*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H72648
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aero*
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72648
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-233 <KAW>
A:Cross-references: DBJ:AP000060; NID:95104188; PIDN:BA079592.1; PID:dl043378; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0622

Query Match 67.9%; Score 36; DB 2; Length 233;
Best Local Similarity 87.5%; Pred. No. 41;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AOPGPVLS 8
 | | | | |
 Db 66 ASPGPVLS 73

RESULT 7
 OOLJX1
 trans-activating transcription regulator - bovine leukemia virus (fragment)
 C:Species: bovine leukemia virus, BLV
 A:Note: host Bos sp. (cattle)
 C>Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 02-Jul-1998
 C:Accession: A04014
 R:Sagata, N.; Yasunaga, T.; Tsuzuku-Kawamura, J.; Ohishi, K.; Ogawa, Y.; Ikawa, Y.
 Proc. Natl. Acad. Sci. U.S.A. 82, 677-681, 1985
 A>Title: Complete nucleotide sequence of the genome of bovine leukemia virus: its evolution
 A:Reference number: A94063; MUID:85140159; PMID:2983308
 A:Accession: A04014
 A:Molecule type: DNA
 A:Residues: 1-308 <SAG>
 C:Superfamily: Leukemia virus trans-activating transcription regulator

Query Match
 Best Local Similarity 67.9%; Score 36; DB 1; Length 308;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AOPGPVLS 10
 | | | | |
 Db 97 AOPGPVLS 106

RESULT 8
 S59531
 manganese peroxidase (EC 1.11.1.13) precursor - oyster mushroom
 C:Species: Pleurotus ostreatus (oyster mushroom)
 C>Date: 20-Jul-1996 #sequence_revision 08-Nov-1996 #text_change 11-Jun-1999
 C:Accession: S59531; S59532; S59525
 R:Asada, Y.; Watanabe, A.; Irie, T.; Nakayama, T.; Kuwahara, M.
 Biochim. Biophys. Acta 1251, 205-209, 1995
 A>Title: Structures of genomic and complementary DNAs coding for Pleurotus ostreatus man
 A:Reference number: S59525; MUID:9539442; PMID:7669812
 A:Accession: S59531
 A:Molecule type: DNA
 A:Residues: 1-361 <ASAS>
 A:Cross-references: EMBL:U21878; NID:g901963; PIDN:AAA84396.1; PID:g726132
 A:Accession: S59532
 A:Molecule type: mRNA
 A:Residues: 1-361 <ASAW>
 A:Cross-references: EMBL:U21879; NID:g732512; PIDN:AAA84397.1; PID:g732513
 A:Accession: S59525
 A:Molecule type: protein
 A:Residues: 30-59 <ASFS>
 C:Genetics:
 A:Insertions: 22/1; 66/3; 86/2; 132/2; 173/1; 198/2; 210/3; 213/2; 240/2; 263/3; 267/1; 279
 C:Superfamily: ligin peroxidase
 C:Keywords: extracellular protein; glycoprotein; heme; manganese; oxidoreductase
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-29/Domain: propeptide #status predicted <PRO>
 F:30-361/Product: manganese (II) peroxidase #status experimental <MAT>

Query Match
 Best Local Similarity 67.9%; Score 36; DB 2; Length 361;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AOPGPVLS 10
 | | | | |
 Db 347 AOPGPVLS 356

RESULT 9
 A75373
 probable N-acetylmutamoyl-L-alanine amidase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: A75373
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
 ; M.; Shen, M.; Yamathayan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: A75373
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-603 <WHIT>
 A:Cross-references: GB:AE002007; GB:AE000513; NID:96459402; PIDN:AAFI1192.1; PID:9645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR1632
 A:Map position: 1

Query Match
 Best Local Similarity 67.9%; Score 36; DB 2; Length 603;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AOPGPVLS 10
 | | | | |
 Db 292 AOPGPVLS 301

RESULT 10
 F82971
 Probable FMN oxidoreductase PA5398 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 02-Nov-2001
 C:Accession: F82971
 R:Stover, C.K.; Pham, X.Q.; Ervin, A.L.; Micoqueli, S.D.; Warren, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardly, K.; L.
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: F82971
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-686 <STO>
 A:Cross-references: GB:AE004952; GB:AE004091; NID:g9951718; PIDN:AGC08783.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA5398
 C:Superfamily: Methylotrophus methylotrophus W3A1 trimethylamine dehydrogenase

Query Match
 Best Local Similarity 67.9%; Score 36; DB 2; Length 686;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 OPGPVL 10
 | | | | |
 Db 646 OPGPVL 654

RESULT 11
 B84062
 hypothetical protein BH3298 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: B84062
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: B84062
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1071 <STO>

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 0.192719 seconds
(Without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53
Sequence: 1 AOPGPVLSQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	801	T2DT_HUMAN	Q92750 homo sapien
2	37	69.8	349	DKK3_MOUSE	Q92750 mus musculu
3	37	69.8	756	IRKB_HUMAN	Q14920 homo sapien
4	37	69.8	1279	SCAP_HUMAN	Q12770 homo sapien
5	36	67.9	308	YPL1_BUY	P03412 bovine leuk
6	36	67.9	1942	Y054_HUMAN	P42694 homo sapien
7	35	66.0	121	SECR_HUMAN	P04683 homo sapien
8	35	66.0	188	COAT_PHV	P36351 physalis mo
9	35	66.0	190	COAT_BMY	P15158 belladonna
10	35	66.0	311	ISPE_MYCLE	Q92750 mycobacteri
11	35	66.0	339	MOXR_PARDE	P29901 paracoccus
12	35	66.0	350	DKK3_HUMAN	Q92750 homo sapien
13	35	66.0	446	PLPD_BACSU	P39130 bacillus su
14	35	66.0	509	NEFA_HUMAN	Q12857 homo sapien
15	35	66.0	509	NEFA_HUMAN	Q12857 homo sapien
16	35	66.0	522	NEFA_RAT	P09414 rattus norv
17	35	66.0	532	NEFA_CHICK	P17923 gallus gall
18	35	66.0	532	NEFA_MOUSE	Q02760 mus musculu
19	35	66.0	587	RGPI_HUMAN	P46060 mus musculu
20	35	66.0	589	RGPI_MOUSE	P46061 mus musculu
21	35	66.0	650	V411_HUMAN	Q43295 homo sapien
22	35	66.0	733	VINE_MOUSE	Q43295 mus musculu
23	34	64.2	110	GON2_SUNMU	O97666 suncus muri
24	34	64.2	236	CYS3_NEUCR	P22697 neurospora
25	34	64.2	298	Y3K_HUMAN	Q04333 homo sapien
26	34	64.2	328	ZIPA_ECO57	Q8X452 escherichia
27	34	64.2	328	ZIPA_ECOLI	P77173 escherichia
28	34	64.2	345	RLAO_PYPAR	Q82436 yersinia pe
29	34	64.2	360	RLAO_PYPAR	Q82436 yersinia pe
30	34	64.2	838	TAC3_HUMAN	O9Y665 homo sapien
31	33	62.3	913	UBPK_HUMAN	O9Y266 homo sapien
32	33	62.3	211	SN23_HUMAN	O00161 homo sapien
33	33	62.3	228	MUTH_ECOLI	P06722 escherichia
33	33	62.3	267	XPA_XENLA	P27088 xenopus lae

34	33	62.3	313	1	CDX2_HUMAN	Q99626 homo sapien
35	33	62.3	354	1	YMA3_MYCO	Q02279 mycobacteri
36	33	62.3	464	1	CDK8_HUMAN	P49336 homo sapien
37	33	62.3	476	1	VL2_HPV32	P36757 human papil
38	33	62.3	531	1	SIS2_CANTR	Q12600 candida tro
39	33	62.3	580	1	SYN3_HUMAN	Q14994 homo sapien
40	33	62.3	603	1	BPRV_BACNO	P42779 bacteroides
41	33	62.3	626	1	PKNH_MYCTU	Q11053 mycobacteri
42	33	62.3	656	1	DNA4_STRCO	P27902 streptomyce
43	33	62.3	890	1	WFS1_HUMAN	P276024 homo sapien
44	33	62.3	890	1	WFS1_MOUSE	P56695 mus musculu
45	33	62.3	1175	1	DSRA_RAT	P5266 rattus norv

ALIGNMENTS

```

RESULT 1
ID T2DT_HUMAN STANDARD: PRT: 801 AA.
AC Q92750;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIID 105 kDa subunit (TAFI1-105)
DE (TAFI1105) (Fragment).
GN TAF1B OR TAF2C2 OR TAF11105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=97011146; PubMed=856156;
RA Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIID subunit related to
RL hTAFII130."
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL-TYPE-SPECIFIC SUBUNIT OF TFIID THAT MAY FUNCTION AS
CC A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIID IS A
CC MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
CC PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
CC HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y09321; CA70499.1; -.
CC Genew: HGNC:11538; TAF1B.
CC MIM: 601689; -.
CC InterPro: IPR003894; TAF_hom.
CC SMART: SM00549; TAFH; 1.
CC Transcription regulation; Nuclear protein.
CC NON_TER 1
CC SEQUENCE 801 AA; 85658 MW; D1B4932FEA49CD2 CRC64;
CC
Query Match 100.0%; Score 53; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
|||||
QY 1 AOPGPVLSQP 10

```

Db 443 AOPGVLSP 452

RESULT 2

DKK3_MOUSE STANDARD; PRT; 349 AA.

AC 090UN9;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE DICKKOPF related protein-3 precursor (DKK-3) (DICKKOPF-3) (mdk-3).

GN DKK3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=9425169; PubMed=10495270;

RA Monaghan P.A., Kioschis P., Wu W., Zuniga A., Bock D., Poustka A.,

RA Dellus H., Niehrs C.;

RT "Dickkopf genes are co-ordinately expressed in mesodermal lineages.";

RL Mech. Dev. 87:45-56(1999).

RN (2)

RP SEQUENCE FROM N.A.

RX MEDLINE=20035735; PubMed=10570958;

RA Krupnik V.E., Sharp J.D., Jiang C., Robison K., Chickering T.W.,

RA Amaratunga L., Brown D.E., Guyot D., Mays G., Leiby K., Chang B.,

RA Duong T., Goodearl A.D.J., Gearing D.P., Sokol S.Y., McCarthy S.A.;

RT "Functional and structural diversity of the human Dickkopf gene family.";

RL Gene 238:301-313(1999).

RN (3)

RP SEQUENCE FROM N.A.

RX STRAIN=C57BL/6J; TISSUE=Liver;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,

RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito K.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleschmann W., Gaasterland T., Glass C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schraml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,

RA Wymshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

CC -1- FUNCTION: INHIBITOR OF WNT SIGNALING PATHWAY (POTENTIAL).

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN, EYE, AND HEART.

CC -1- SIMILARITY: BELONGS TO THE DICKKOPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: AJ243964; CAB60111.1; -

DR EMBL: AF177400; AA02680.1; -

DR EMBL: AK004853; BAB23617.1; -

MD: MGI:1354952; DKK3.

KW Developmental protein; Signal; Glycoprotein.

FT SIGNAL. 1 22

FT CHAIN 23 349

FT DOMAIN 147 195

FT DOMAIN 208 284

FT CARBOHYD 96 96

FT CARBOHYD 106 106

FT CARBOHYD 121 121

FT CARBOHYD 204 204

SO SEQUENCE 349 AA; 38387 MW; 564CB3C4FB2EA88 CRC64;

Query Match 69.8%; Score 37; DB 1; Length 349;

Best Local Similarity 60.0%; Pred. No. 18;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AOPGVLSP 10

Db 32 AOPGVLSP 41

RESULT 3

IKKB_HUMAN STANDARD; PRT; 756 AA.

AC 014920; 075327;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Inhibitor of nuclear factor kappa B kinase beta subunit (IC 2.7.1.-)

DE (1-kappa-B-kinase beta) (IKKB) (IKK-B) (I-kappa-B kinase

DE 2) (IKK2) (Nuclear factor NF-kappa-B inhibitor kinase beta) (NFKB1KB).

GN IKKB OR IKKBK.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A., AND MUTAGENESIS OF LYS-44; SER-177 AND SER-181.

RX TISSUE=Cervical carcinoma;

RX MEDLINE=98008813; PubMed=9346484;

RA Mercuro F., Zhu H., Murray B.W., Shevchenko A., Bennett B.L.,

RA Li J.W., Young D.B., Barbosa M., Mann M., Manning A., Rao A.;

RT "IKK-1 and IKK-2: cytochrome-activated IkappaB kinases essential for NF-kappaB activation.";

RL Science 278:860-866(1997).

RN (2)

RP SEQUENCE FROM N.A., AND MUTAGENESIS OF LYS-44.

RX MEDLINE=98008814; PubMed=9346485;

RA Moronicz J.D., Gao X., Cao Z., Rothe M., Goeddel D.V.;

RT "IkappaB kinase-beta: NF-kappaB activation and complex formation with IkappaB kinase-alpha and NIK.";

RL Science 278:866-869(1997).

RN (3)

RP SEQUENCE FROM N.A.

RX TISSUE=Heart;

RX MEDLINE=99032998; PubMed=9813230;

RA Hu M.C.-T., Wang Y.-P.;

RT "IkappaB kinase alpha and -beta genes are coexpressed in adult and embryonic tissues but localized to different human chromosomes.";

RL Gene 223:31-40(1998).

RN (4)

RP SEQUENCE FROM N.A., AND GENE MAPPING.

RX MEDLINE=98438415; PubMed=9763654;

RA Shindo M., Nakano H., Sakon S., Yagita H., Mihara M., Okumura K.;

RT "Assignment of IkappaB kinase beta (IKKB) to human chromosome band 8p12--p11 by in situ hybridization.";

RL Cytogenet. Cell Genet. 82:32-33(1998).

RN (5)

RP SEQUENCE OF 1-256 FROM N.A.

RX TISSUE=Lung;

RA Stransberg R.;

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

[6]

IKK PHOSPHORYLATION.

RX MEDLINE-99038238; PubMed-9819420;
 RA Nemoto S., Didonato J.A., Lin A.;
 RT "Coordinate regulation of IkappaB kinases by mitogen-activated protein
 RT kinase kinase kinase 1 and NF-kappaB-inducing kinase.";
 RL Mol. Cell. Biol. 18:7336-7343(1998).
 RN [7]
 RP REVIEW.
 RX MEDLINE-20178139; PubMed-10712233;
 RA Jobin C., Sartor R.B.;
 RT "The I kappa B/NF-kappa B system: a key determinant of
 RT mucosal inflammation and protection.";
 RL Am. J. Physiol. 278:C451-C462(2000).
 CC -1- FUNCTION: PHOSPHORYLATES INHIBITORS OF NF-KAPPA-B THUS LEADING TO
 CC THE DISSOCIATION OF THE INHIBITOR/NF-KAPPA-B COMPLEX AND
 CC ULTIMATELY THE DEGRADATION OF THE INHIBITOR.
 CC -1- SUBUNIT: PREFERENTIALLY FOUND AS A HETERODIMER WITH IKK-ALPHA BUT
 CC ALSO AS A HOMODIMER. DIRECTLY INTERACTS WITH IKK-GAMMA/NEMO.
 CC HETERODIMERS FORM THE ACTIVE COMPLEX. THE TRIPARTITE COMPLEX CAN
 CC ALSO BIND TO MEK1, MAPK14/NIK, IKAP AND IKB-ALPHA-P65-P50
 CC COMPLEX. PHOSPHORYLATED IKB-ALPHA IS FURTHER RELEASED FROM THE
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, PLACENTA, SKELETAL
 CC MUSCLE, KIDNEY, PANCREAS, SPLEEN, THYMOS, PROSTATE, TESTIS AND
 CC PERIPHERAL BLOOD.
 CC -1- PTM: PHOSPHORYLATED BY MEK1 AND PROBABLY ALSO BY MAPK14/NIK.
 CC WEAKLY AUTOPHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC IKAPPAB KINASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF029684; AAC51860.1; -;
 DR EMBL: AF080158; AAD08997.1; -;
 DR EMBL: AF031416; AAC64675.1; -;
 DR EMBL: BC006231; AAH06231.1; -;
 DR HSSP: Q63450; IA06.
 DR Genew: HGNC:5960; IKKB.
 DR MIM: 603258; -;
 DR InterPro: IPR000719; Euk.pkinase.
 DR InterPro: IPR002290; Ser-thr.pkinase.
 DR InterPro: IPR001245; Tyr.pkinase.
 DR Pfam: PF00069; pkinase.1.
 DR Pfam: PF00240; ubiqtulin.1.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk.pkinase.1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 DR Translase: Serine/threonine-protein kinase; ATP-binding;
 DR phosphorylation.
 KW DOMAIN 15 300 PROTEIN KINASE.
 FT DOMAIN 15 300 LEUCINE-ZIPPER (POTENTIAL).
 FT DOMAIN 458 479 NEMO-BINDING.
 FT NP_BIND 737 742 ATP (BY SIMILARITY).
 FT BINDING 21 29 ATP (BY SIMILARITY).
 FT ACT_SITE 44 44 ATP (BY SIMILARITY).
 FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 177 177 PHOSPHORYLATION.
 FT MOD_RES 181 181 PHOSPHORYLATION.
 FT MUTAGEN 44 44 K->A: LOSS OF KINASE ACTIVITY AND NO
 FT MUTAGEN 177 177 EFFECT ON BINDING TO NIK.
 FT MUTAGEN 177 177 S->A: DECREASE OF ACTIVITY.
 FT MUTAGEN 181 181 S->E: FULL ACTIVATION.
 FT MUTAGEN 181 181 S->A: DECREASE OF ACTIVITY.
 FT CONFLICT 231 255 WSKVRKSEVDIVSEDLNCTVKE -> CVRMMPGVVAHS

FT CONFLICT 425 425 CNPSTLGGRGHWI (IN REF. 5).
 FT SEQUENCE 756 AA; 86563 MW; F9CADF671AE9E1AE CRC64;
 SQ
 Query Match 69.8%; Score 37; DB 1; Length 756;
 Best Local Similarity 60.0%; Pred. No. 41;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 AQPGLVLSQP 10
 Db 682 SQPGLMSQP 691
 RESULT 4
 SCAP_HUMAN STANDARD; PRT; 1279 AA.
 AC Q12770;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sterol regulatory element binding protein cleavage-activating protein
 DE (SREBP cleavage-activating protein) (SCAP).
 DE SCAP OR KIAA0199.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Hepatoma;
 RA Nakajima T., Hama Kubo T., Kodama T., Inazawa J., Eml M.;
 RT MEDLINE-20037503; PubMed-10570913;
 RT "Genomic structure and chromosomal mapping of the human sterol
 RT regulatory element binding protein (SREBP) cleavage-activating
 RT protein (SCAP) gene";
 RL J. Hum. Genet. 44:402-407(1999).
 RN [2]
 RP SEQUENCE OF 3-1279 FROM N.A.
 RC TISSUE-Bone marrow;
 RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
 RT MEDLINE-96281124; PubMed-8724849;
 RT "Prediction of the coding sequences of unidentified human genes. V.
 RT The coding sequences of 40 new genes (K1AA0161-K1AA0200) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 3:17-24(1996).
 RN [3]
 RP SEQUENCE OF 346-1279 FROM N.A.
 RC TISSUE-Teratocarcinoma;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Magatsuma M., Hosokari T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
 RA Minomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP VARIANT ILE-798.
 RX MEDLINE-20037509; PubMed-10570919;
 RA Iwaki K., Nakajima T., Ota N., Eml M.;
 RT "A common Ile796Val polymorphism of the human SREBP cleavage-
 RT activating protein (SCAP) gene.";
 RL J. Hum. Genet. 44:421-422(1999).
 CC -1- FUNCTION: Sterol sensor. Necessary for the proteolytic activation
 CC of SREBPs by site-1 protease in the Golgi.
 CC -1- SUBUNIT: In a tight complex with SREBP (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Moves from the
 CC endoplasmic reticulum to the Golgi in the absence of sterols (By
 CC similarity).
 CC -1- SIMILARITY: CONTAINS 1 STEROL SENSOR DOMAIN (SSD).
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D83782; BAA12111.1; -
 DR EMBL: AK027402; BAB5088.1; ALT_INIT.
 DR MIM: 601510; -
 DR InterPro: IPR000731; HMGCR/Patch_5TM.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 12.
 DR PRINTS: PR00320; GPROTEINBRPT.
 DR SMART: SM00320; WD40; 10.
 DR PROSITE: PS00156; SSD; 1.
 DR PROSITE: PS00082; WD_REPEATS_2; 1.
 DR PROSITE: PS00294; WD_REPEATS_REGION; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; 1.
 KW Lipid metabolism; Cholesterol metabolism; Transmembrane;
 KW Endoplasmic reticulum; Golgi stack; Repeat; WD repeat; Glycoprotein;
 KW Polymorphism.
 FT DOMAIN 1 18
 FT TRANSMEM 19 39
 FT DOMAIN 40 279
 FT TRANSMEM 280 300
 FT DOMAIN 301 312
 FT TRANSMEM 313 333
 FT DOMAIN 334 344
 FT TRANSMEM 345 365
 FT DOMAIN 366 401
 FT TRANSMEM 402 422
 FT DOMAIN 423 423
 FT TRANSMEM 424 444
 FT DOMAIN 445 518
 FT TRANSMEM 519 539
 FT DOMAIN 540 709
 FT TRANSMEM 710 730
 FT DOMAIN 731 1279
 FT TRANSMEM 284 442
 FT DOMAIN 771 811
 FT REPEAT 952 1002
 FT REPEAT 1005 1042
 FT REPEAT 1077 1114
 FT REPEAT 1117 1155
 FT REPEAT 1158 1195
 FT REPEAT 1197 1235
 FT DOMAIN 722 729
 FT DOMAIN 747 750
 FT DOMAIN 861 864
 FT CARBOHYD 263 263
 FT CARBOHYD 590 590
 FT CARBOHYD 641 641
 FT CARBOHYD 798 798
 FT VARIANT
 FT CONFLICT 617 617
 FT CONFLICT 753 753
 FT CONFLICT 817 817
 FT CONFLICT 941 941
 FT CONFLICT 1035 1035
 FT CONFLICT 1080 1080
 FT CONFLICT 1279 AA; 139728 MW; 4CD73543B7D2ACB4 CRC64;
 SO SEQUENCE

Query Match 69.8%; Score 37; DB 1; Length 1279;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 PGPVLSQ 9
 DB 938 PGPVLSQ 944

RESULT 5
 ID YPX1_BLVU STANDARD; PRT; 308 AA.
 AC P03412;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE Hypothetical PXL-I protein (fragment).
 OS Bovine leukemia virus (Japanese isolate BLV-1) (BLV).
 OC Viruses; Retroviridae; Retroviridae; Deltaretrovirus.
 OX NCBI_Taxid=11907;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=65140159; PubMed=2983308;
 RA Segata N., Yasunaga T., Tsuzuku-Kawamura J., Ohishi K., Ogawa Y.,
 RA Ikawa Y.;
 RT Complete nucleotide sequence of the genome of bovine leukemia virus:
 RT its evolutionary relationship to other retroviruses.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:677-681(1985).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: K02120; -; NOT_ANNOTATED_CDS.
 DR PIR: A04014; GOLLX1.
 KW Hypothetical protein.
 FT NON_TER 1
 FT SEQUENCE 308 AA; 34191 MW; 082AD897E614672B CRC64;
 SO SEQUENCE

Query Match 67.9%; Score 36; DB 1; Length 308;
 Best Local Similarity 70.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AOPGVLSQ 10
 DB 97 AOPGVLSQ 106

RESULT 6
 ID Y054_HUMAN STANDARD; PRT; 1942 AA.
 AC P42694;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein KIAA0054.
 DE KIAA0054.
 GN KIAA0054.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Bone marrow;
 RX MEDLINE=96051398; PubMed=7584044;
 RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
 RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
 RT Prediction of the coding sequences of unidentified human genes. II.
 RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
 RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 1:223-229(1994).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D29677; BAA06147.1; -

DR Genew: HGNC:16878; HELZ.

DR InterPro: IPR000571; Znf_CCH.

DR Pfam: PF00642; Zf-CCH1.1.

DR SMART: SM00356; Znf_CCH1.1.

KW Hypothetical protein; Zinc-finger.

SO SEQUENCE 1942 AA; 218970 MW; 94D13601A02335E0 CRC64;

Query Match 67.9%; Score 36; DB 1; Length 1942;

Best Local Similarity 85.7%; Pred. No. 1.7e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AOPGPVL 7

Db 1469 AOPGPVL 1475

RESULT 7

SECR_HUMAN STANDARD; PRT; 121 AA.

AC P09683;

DT 01-MAR-1989 (Rel. 10, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Secretin precursor.

GN SCT.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20515579; PubMed=11060443;

RA Whitmore T.E., Holloway J.L., Lofton-Day C.E., Maurer M.F., Chen L., Quinton T.J., Vincent J.B., Scherer S.W., Lok S.;

RT "Human secretin (SCT): gene structure, chromosome location, and distribution of mRNA".

RT Cytogenet. Cell Genet. 90:47-52(2000).

RL [2]

RN SEQUENCE OF 28-54.

RA Carlquist M., Joernvall H., Forssmann W.-G., Thulin L., Johansson C., Mutt V.;

RT "Human secretin is not identical to the porcine/bovine hormone.";

RL ICS Med. Sci. 13:217-218(1985).

CC -1- FUNCTION: STIMULATES FORMATION OF NAHCO(3)-RICH PANCREATIC JUICE AND SECRETION OF NAHCO(3)-RICH BILE AND INHIBITS HCL PRODUCTION BY THE STOMACH.

CC -1- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF244355; AAC31443.1; -

DR PIR: S07443; S07443.

DR Genew: HGNC:10607; SCT.

DR MIM: 182099; -

DR InterPro: IPR000532; Glucagon.

DR Pfam: PF00123; hormone2; 1.

DR SMART: SM0070; GLUCA; 1.

DR PROSITE: PS00260; GLUCAGON; 1.

KW Glucagon family; Hormone; Amidation; Cleavage on pair of basic residues; Signal.

FT SIGNAL 1

FT PEPTIDE 18

FT MOD_RES 34 54 SECRETIN. AMIDATION (G-55 PROVIDE AMIDE GROUP).

SO SEQUENCE 121 AA; 13016 MW; 44BD4EFC0E161CF CRC64;

Query Match 66.0%; Score 35; DB 1; Length 121;

Best Local Similarity 62.5%; Pred. No. 13;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 3 GPGVLSQP 10

Db 102 GPGVLSQP 109

RESULT 8

COAT_PHV STANDARD; PRT; 188 AA.

ID COAT_PHV

AC P36351;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE Coat protein (Virion protein).

OS Physalis mottle virus (PhMV) (Belladonna mottle virus-Iowa).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviruses.

OX NCBI_Taxid=72539;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9222369; PubMed=1562236;

RA Jacob A.N.K., Murthy M.R., Savithri H.S.;

RT "Nucleotide sequence of the 3' terminal region of belladonna mottle virus-Iowa (renamed Physalis mottle virus) RNA and an analysis of the relationships of tymoviral coat proteins.";

RT Arch. Virol. 123:367-377(1992).

RL [2]

RN SEQUENCE FROM N.A.

RX MEDLINE=93212529; PubMed=8460497;

RA Kekuda R., Karande A.A., Jacob A.N.K., Savithri H.S.;

RT "Architecture of physalis mottle tymovirus as probed by monoclonal antibodies and cross-linking studies.";

RL Virology 193:959-966(1993).

CC -1- SUBUNIT: THE VIRUS COAT IS COMPOSED OF 180 COPIES OF THE COAT PROTEIN ARRANGED IN A ICOSAEDRAL SHELL.

CC -1- SIMILARITY: TO COAT PROTEIN FROM OTHER TYMOVIRUSES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: S97776; AAB21997.1; -

DR PIR: B45540; B45540.

DR HSSP: P03608; IADV.

DR InterPro: IPR000574; Tymo_coat.

DR Pfam: PF00983; Tymo_coat; 1.

KW Coat protein.

SO SEQUENCE 188 AA; 19974 MW; DIDC64831FE5DC1 CRC64;

Query Match 66.0%; Score 35; DB 1; Length 188;

Best Local Similarity 75.0%; Pred. No. 21;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 3 GPGVLSQP 10

Db 17 PGSLISQP 24

RESULT 9

COAT_BMDV STANDARD; PRT; 190 AA.

ID COAT_BMDV

AC P15158;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

CC	COAT PROTEIN (VIRION PROTEIN).	66.0%;	Score 35;	DB 1;	Length 190;
OS	Belladonna mottle virus (BMDV).	Best Local Similarity	75.0%;	Pred. No. 22;	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.	Matches	6;	Conservative	1;
ON	NB1_taxid=12149;	Mismatches	1;	Indels	0;
XX	[1]	Gaps			
RP	SEQUENCE.				
RX	MEDLINE=69197927; PubMed=2467911;				
RA	Subviral S., Rao N.A., Murthy M.R.N., Savithri H.S.;				
RT	"Primary structure of belladonna mottle virus coat protein.";				
RL	J. Biol. Chem. 264:6273-6279(1989).				
RN	[2]				
RP	SEQUENCE OF 144-186.				
RA	Subviral S., Rao D.R., Rao N.A., Savithri H.S.;				
RT	"Partial amino acid sequence of belladonna mottle virus coat protein.";				
RL	Biochem. Int. 15:65-72(1987).				
CC	-1- SUBUNIT: THE VIRUS COAT IS COMPOSED OF 180 COPIES OF THE COAT				
CC	PROTEIN ARRANGED IN AN ICOSAEDRAL SHELL.				
CC	-1- SIMILARITY: TO COAT PROTEIN FROM OTHER TYMOVIRUSES.				
DR	PIR: S02207; S02207.				
DR	HSP: P03608; 1AU.				
DR	InterPro: IPR00574; Tymo_coat.				
DR	Pfam: PF00983; Tymo_coat; 1.				
KW	Coat protein; Acetylation.				
FT	MOD_RES 1 ACETYLATION.				
FT	CONFLICT 179 179 R -> K (IN REF. 2)				
FT	CONFLICT 184 186 KPM -> SPQ (IN REF. 2)				
SO	SEQUENCE 190 AA; 20215 MW; 20EEB23DD863E14 CRC64;				
OY	3 PGVLSQP 10				
Db	11:1111				
	17 PGSLISQP 24				
RESULT 10					
ISPE_MYCLE	STANDARD; PRT; 311 AA.				
ID	ISPE_MYCLE				
AC	Q9CD51;				
DT	15-JUN-2002 (Rel. 41, Created)				
DT	15-JUN-2002 (Rel. 41, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK				
DE	4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).				
DE	ISPE OR ML0242.				
OS	Mycobacterium leprae.				
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;				
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.				
OX	NB1_taxid=1769;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
CC	STRAIN-TN:				
RX	MEDLINE=21128732; PubMed=11234002;				
RA	Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,				
RA	Wheeler P.R., Honore K., Garnier T., Churcher C., Harris D.,				
RA	Mungall K., Baahm D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,				
RA	Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean A., Hamlin N.,				
RA	Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,				
RA	Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,				
RA	Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,				
RA	Barrell B.G.;				
RT	"Massive gene decay in the leprosy bacillus.";				
RL	Nature 409:1007-1011(2001).				
CC	-1- FUNCTION: Catalyses the phosphorylation of the position 2 hydroxy				
CC	group of 4-diphosphocytidyl-2C-methyl-D-erythritol (by				
CC	similarity).				
CC	-1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-				
CC	erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-				
CC	methyl-D-erythritol.				

Search completed: February 16, 2003, 21:56:23
 Job time : 2.19272 secs

GN NFIA OR NF1L21.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Liver;
 RX MEDLINE=96429753; PubMed=8832903;
 RA Monaci P., Nuzzo M., Stampfl S., Tollervey D., de Simone V.,
 RA Nicotia A.;
 RT "A complex interplay of positive and negative elements is responsible
 for the different transcriptional activity of liver NF1 variants.";
 RL Mol. Biol. Rep. 21:147-158(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Liver;
 RX MEDLINE=97244636; PubMed=9089412;
 RA Osada S., Daimon S., Ikeda T., Nishihara T., Yano K., Yamasaki M.,
 RA Imagawa M.;
 RT "Nuclear factor 1 family proteins bind to the silencer element in the
 rat glutathione transferase P gene.";
 RL J. Biochem. 121:355-363(1997).
 RN [3]
 RP SEQUENCE OF 5-509 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Liver;
 RX MEDLINE=09030607; PubMed=3053160;
 RA Paoessa G., Gonnari F., Frank R., Cortese R.;
 RT "Purification of a NF1-like DNA-binding protein from rat liver and
 cloning of the corresponding cDNA.";
 RL EMBO J. 7:3115-3123(1988).
 CC -1- FUNCTION: RECOGNIZES AND BINDS THE PALINDROMIC SEQUENCE 5'-
 TTGGCANNNGCCAA-3' PRESENT IN VIRAL AND CELLULAR PROMOTERS AND IN
 INDIVIDUALLY CAPABLE OF ACTIVATING TRANSCRIPTION AND REPLICATION.
 CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE CTF/NF-1 FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 DNA/MH1 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X84209; CAA58995.1; -;
 DR EMBL: D78017; BAAL1203.1; -;
 DR EMBL: X13167; CAA31565.1; -;
 DR PIR: S01300; S01300.
 DR TRANSFAC: T00535; -;
 DR TRANSFAC: T00589; -;
 DR TRANSFAC: T02296; -;
 DR InterPro: IPR000647; CTF_NF1_family.
 DR InterPro: IPR003619; Dwarfin_A.
 DR Pfam: PF00859; CTF_NF1; 1.
 DR SMART: SM00523; DNA; 1.
 DR PROSITE: PS00349; CTF_NF1; 1.
 KW Transcription regulation; DNA replication; DNA-binding; Activator;
 KW Nuclear protein; Multigene family.
 FT DOMAIN 67 175 DNA.
 SO SEQUENCE 509 AA; 55976 MW; 59120C7090229F87 CRC64;

Query Match 66.0%; Score 35; DB 1; Length 509;
 Best Local Similarity 66.7%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 OPGPVLSP 10
 DB 378 OPGPFLSHP 386

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 0.702355 Seconds
(without alignments)
2933.659 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_452

Perfect score: 53
Sequence: 1 AQPcPVLISQP 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rv1rus:*
16: sp_bacteriapi:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	77.4	354	10	041328 lycopersico
2	41	77.4	360	10	091KY4 glycine max
3	41	77.4	360	10	091KY3 glycine max
4	41	77.4	364	10	091UT0 arabidopsis
5	41	77.4	365	10	080719 arabidopsis
6	39	73.6	258	5	090191 leishmania
7	39	73.6	797	4	09UGS3 homo sapien
8	38	71.7	244	4	08WY22 homo sapien
9	38	71.7	265	2	093G77 streptomyce
10	38	71.7	397	4	096TW9 homo sapien
11	38	71.7	397	4	096TW9 homo sapien
12	38	71.7	583	16	08XUK6 raistonia s
13	38	71.7	583	16	08XUK6 raistonia s
14	38	71.7	649	11	099L19 mus musculu
15	38	71.7	728	4	098T94 homo sapien
16	38	71.7	818	4	09C0B0 homo sapien

17	38	71.7	824	4	08TF48	08TF48 homo sapien
18	38	71.7	942	4	096K70	096K70 homo sapien
19	38	71.7	987	16	09RWP6	09RWP6 delnocoocus
20	37	69.8	349	11	09C231	09C231 mus musculu
21	37	69.8	349	16	09HV34	09HV34 pseudomonas
22	37	69.8	371	10	09AMP9	09AMP9 oryza sativ
23	37	69.8	465	5	08WS46	08WS46 oikopleura
24	37	69.8	553	6	095KH7	095KH7 macaca fasc
25	37	69.8	595	10	08S1W9	08S1W9 oryza sativ
26	37	69.8	891	5	093763	093763 caenorhabdl
27	37	69.8	905	4	08WU41	08WU41 homo sapien
28	36	67.9	100	6	077624	077624 bos taurus
29	36	67.9	163	16	09PE02	09PE02 xylella fas
30	36	67.9	180	16	08XRY8	08XRY8 raistonia s
31	36	67.9	233	17	09YEF4	09YEF4 aeropyrum p
32	36	67.9	309	15	091G01	091G01 bovine leuk
33	36	67.9	309	15	085493	085493 bovine leuk
34	36	67.9	309	15	056232	056232 pleurotus o
35	36	67.9	361	3	012332	012332 pleurotus o
36	36	67.9	361	3	094753	094753 pleurotus e
37	36	67.9	361	3	09UR19	09UR19 pleurotus e
38	36	67.9	362	13	092074	092074 gallus gall
39	36	67.9	374	4	09NS36	09NS36 homo sapien
40	36	67.9	511	16	093RY1	093RY1 streptomyce
41	36	67.9	580	2	09L705	09L705 kylococcus
42	36	67.9	592	11	097301	097301 mus musculu
43	36	67.9	603	16	09RTX0	09RTX0 delnocoocus
44	36	67.9	686	16	09HMC6	09HMC6 pseudomonas
45	36	67.9	867	4	09NU50	09NU50 homo sapien

ALIGNMENTS

RESULT 1
ID 041328 PRELIMINARY; PRT; 354 AA.
AC 041328;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Pto Kinase Interactor 1.
GN P11.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO GRANDE-PTOR;
RX MEDLINE=96107331; PubMed=8521516;
RA Zhou J., Loh Y.T., Bressan R.A., Martin G.B.;
RT "The tomato gene P11 encodes a serine/threonine kinase that is
RT phosphorylated by Pto and is involved in the hypersensitive
RT response.";
RL Cell 83:925-935(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO GRANDE-PTOR;
RA Martin G.B.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: U28007; AAC61805.1; -;
DR InterPro: IPR00719; Euk_pkinase.
DR InterPro: IPR002203; Intein.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR004040; STY_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00221; STYK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PSS0011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS00881; PROTEIN_SPLICING; UNKNOWN_1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 354 AA; 38540 MW; A6C5346175FCBE97 CRC64;
Query Match 77.4%; Score 41; DB 10; Length 354;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLS 8
| | | | | | | |
Db 161 AOPGPVLS 168
RESULT 2
O9LUK4 PRELIMINARY: PRT; 360 AA.
ID O9LUK4
AC O9LUK4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE PtlI kinase-like protein.
GN PtlIA.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21036115; PubMed-11185558;
RA Staswick P.;
RT "Two expressed soybean genes with high sequence identity to tomato
PtlI kinase lack autophosphorylation activity.";
RL Arch. Biochem. Biophys. 383:233-237(2000).
DR EMBL; AF249317; AAF91336.1; -
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00221; STYK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 360 AA; 39375 MW; E170D95DD20E2D6F CRC64;
Query Match 77.4%; Score 41; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLS 8
| | | | | | | |
Db 160 AOPGPVLS 167
RESULT 3
O9LUK3 PRELIMINARY: PRT; 360 AA.
ID O9LUK3
AC O9LUK3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE PtlI kinase-like protein.
GN PtlIB.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE-21036115; PubMed-11185558;
RA Staswick P.;
RT "Two expressed soybean genes with high sequence identity to tomato
PtlI kinase lack autophosphorylation activity.";
RL Arch. Biochem. Biophys. 383:233-237(2000).
DR EMBL; AF249318; AAF91337.1; -
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00221; STYK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 360 AA; 39481 MW; 0B54143D95C3B90 CRC64;
Query Match 77.4%; Score 41; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AOPGPVLS 8
| | | | | | | |
Db 160 AOPGPVLS 167
RESULT 4
O9LUT0 PRELIMINARY: PRT; 364 AA.
ID O9LUT0
AC O9LUT0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE serine/threonine protein kinase-like protein (Hypothetical 39.6 kDa
protein).
GN At3g17410.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Nakamura Y.;
RL "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty P1 and TAC
clones.";
RL DNA Res. 7:131-135(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
Lee J.M., Ondera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
Yamamura Y., Yu G., Yu S., Bowser L., Carlinici P., Chen H., Cheuk R.,
Hayashizaki Y., Ishida J., Jones T., Kamuya A., Karlin-Neumann G.,
Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
Theologis A.;
RT "Full length cDNA of gene At3g17410 (GI:15229035).";
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AB022216; BAB02745.1; -
DR EMBL; AY070413; AAL49909.1; -
DR HSSP; P11362; IFGI
InterPro; IPR000719; Euk_pkinase.

DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR004040; STY_pkinase.
DR InterPro: IPR001245; TYR_pkinase.
DR Pfam: PF00069; pkinase.1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase.1.
DR SMART: SM00221; STYKc.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding: Kinase: Serine/threonine-protein kinase; Transferase;
KW Hypothetical protein
SQ SEQUENCE 364 AA; 39562 MW; 62A7FE8F249F332E CRC64;

Query Match 77.4%; Score 41; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQP6PVL5 8
DB 162 AQP6PVL5 169

RESULT 5
080719 PRELIMINARY; PRT; 365 AA.
AC 080719;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative pto kinase (Hypothetical 39.9 kDa protein).
GN F14M.11 OR AT2G47060.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.,
RT "Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Theologis A., Ecker J., Davis R.W.,
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Arabidopsis full length cDNA clones.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AC004411; AAC34243.1; -;
DR EMBL: AF325090; AK17158.1; -;
DR EMBL: AY080876; AAL87347.1; -;
DR HSSP: P11362; IFGI.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR004040; STY_pkinase.
DR Pfam: PF00069; pkinase.1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00221; STYKc.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein; ATP-binding; Serine/threonine-protein kinase;
KW Transferase.
SQ SEQUENCE 365 AA; 39931 MW; ACAFECC6C79E77CAB CRC64;

Query Match 77.4%; Score 41; DB 10; Length 365;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQP6PVL5 8
DB 166 AQP6PVL5 173

RESULT 6
090191 PRELIMINARY; PRT; 258 AA.
AC 090191;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical transmembrane protein L302.06.
GN L302.06.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIDLIN;
RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
RA Rajandream M.A., Barrell B.G.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIDLIN;
RA MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL: AL389894; CAC22674.1; -;
KW Transmembrane.
SQ SEQUENCE 258 AA; 27729 MW; 1F214B2109E0D2DB CRC64;

Query Match 73.6%; Score 39; DB 5; Length 258;
Best Local Similarity 70.0%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AQP6PVL5OP 10
DB 211 AQP6PVL5SP 220

RESULT 7
090G53 PRELIMINARY; PRT; 797 AA.
AC 090G53;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE D1756G23.1 (Novel leucine rich protein) (Fragment).
GN D1756G23.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Corby N.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035681; CAB3072.1; -;
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.

DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003591; LRR_cyp.
 DR Pfam: PF00560; LRR_18.
 DR Pfam: PF01463; LRRCT_2.
 DR Pfam: PF01462; LRRNT_1.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00370; LRR_8.
 DR SMART: SM00082; LRRCT_2.
 DR SMART: SM00013; LRRNT_2.
 DR SMART: SM00369; LRR_TYP_4.
 DR NON_TER 797
 FT SEQUENCE 797 AA; 85614 MW; 8C3247883EAE59AD CRC64;

Query Match 73.6%; Score 39; DB 4; Length 797;
 Best Local Similarity 88.9%; Pred. No. 59;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AOPGVLSQ 9
 DB 261 ALPGVLSQ 269

RESULT 8

ID 08WV2 PRELIMINARY; PRT; 244 AA.
 AC 08WV2: 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical 27.2 kDa protein.
 GN P14776.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang P.P., Zhou X.M., Jiang H.Q., Huang Y., Qin W.X., Zhao X.T.,
 Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF18347; AAL5854.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 244 AA; 27187 MW; 5E7520BB8245F8AB CRC64;

Query Match 71.7%; Score 38; DB 4; Length 244;
 Best Local Similarity 66.7%; Pred. No. 28;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 OPGVLSQ 10
 DB 209 OPGVLSQ 217

RESULT 9

ID 093GV7 PRELIMINARY; PRT; 265 AA.
 AC 093GV7:
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Thioesterase.
 OS Streptomyces avermilitis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,

RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermilitis: Deducing the ability of producing secondary
 RT metabolites."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 DR EMBL: AB070956; BAB69404.1; -
 DR InterPro: IPR001031; Thioesterase.
 DR Pfam: PF00975; Thioesterase; 1.
 SQ SEQUENCE 265 AA; 28688 MW; 70DB99578E56C48A CRC64;

Query Match 71.7%; Score 38; DB 2; Length 265;
 Best Local Similarity 87.5%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PGVLSQ 10
 DB 194 PGVLSQ 201

RESULT 10

ID 096JW9 PRELIMINARY; PRT; 397 AA.
 AC 096JW9:
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE CDNA FLJ14925 f1s, clone PLACE1008643, weakly similar to
 DE Inter-alpha-trypsin inhibitor heavy chain H2 precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murekami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
 RA Niomiya K., Iwayanagi T.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK027831; BAB55397.1; -
 DR InterPro: IPR002035; VWF_A.
 DR Pfam: PF00092; vwa; 1.
 DR PROSITE: PS50234; VWF_A; 1.
 SQ SEQUENCE 397 AA; 44345 MW; 096673471189AF75 CRC64;

Query Match 71.7%; Score 38; DB 4; Length 397;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 OPGVLSQ 10
 DB 353 OPGVLSQ 361

RESULT 11

ID 096JW8 PRELIMINARY; PRT; 397 AA.
 AC 096JW8:
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CDNA FLJ14943 f1s, clone PLACE1011371, weakly similar to
 DE Inter-alpha-trypsin inhibitor heavy chain H2 precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.,
RT "NDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027849; BAB55409.1; -.
DR InterPro: IPR002035; WVF_A.
DR Pfam: PF00092; WVF.1.
DR PROSITE: PS50234; WVF.A.1.
SQ SEQUENCE 397 AA; 44342 MW; 9B41AA56E65AF64 CRC64;

Query Match 71.7%; Score 38; DB 4; Length 397;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QPGPVLSQP 10
Db 353 QPGPLTKRP 361

RESULT 12
Q8XUK6 PRELIMINARY; PRT; 583 AA.
ID 08XUK6
AC Q8XUK6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Probable activation/secretion signal peptide protein.
GN RSC3182 OR RS06080.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
CX Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RX MEDLINE-21681879; PubMed-11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
RA Ariat M., Billault A., Brotlier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebault P., Whalen M., Winkler P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RL "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
RW EMBL: AL646074; CAD16970.1; -.
KW Complete proteome.
SQ SEQUENCE 583 AA; 62313 MW; CEDE73F9943A5573 CRC64;

Query Match 71.7%; Score 38; DB 16; Length 583;
Best Local Similarity 70.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
Db 26 AOPGPVLEDP 35

RESULT 13
Q8XPV1 PRELIMINARY; PRT; 583 AA.
ID 08XPV1
AC Q8XPV1;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Probable activation/secretion signal peptide protein.
GN RSP1535 OR RS06081.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
CX Plasmid megaplasmid.

```

```

OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RX MEDLINE-21681879; PubMed-11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
RA Ariat M., Billault A., Brotlier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebault P., Whalen M., Winkler P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RL "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
RW EMBL: AL646085; CAD18686.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 583 AA; 62554 MW; C248BDA5001DF3C3 CRC64;

Query Match 71.7%; Score 38; DB 16; Length 583;
Best Local Similarity 70.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AOPGPVLSQP 10
Db 26 AOPGPVLEDP 35

RESULT 14
Q99LJ9 PRELIMINARY; PRT; 649 AA.
ID Q99LJ9
AC Q99LJ9;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 69.6 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC003195; AAH03195.1; -.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR000571; Znf_CCH.
DR InterPro: IPR001841; Znf_CCH.
DR Pfam: PF00642; Zf-CCH; 2.
DR SMART: SM00184; RING; 1.
DR SMART: SM00356; Znf_C3H1; 3.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 649 AA; 69599 MW; C6B161FB791D2CEB CRC64;

Query Match 71.7%; Score 38; DB 11; Length 649;
Best Local Similarity 77.8%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 QPGPVLSQP 10
Db 178 QPGPVLYMP 186

RESULT 15
Q9BT94 PRELIMINARY; PRT; 728 AA.
ID Q9BT94
AC Q9BT94;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 80.6 kDa protein.

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004282; AAH04282.1; -
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWF_A; 1.
KW Hypothetical protein.
SQ SEQUENCE 728 AA; 80604 MW; 118A1B32F89BCECA CRC64;

Query Match 71.7%; Score 38; DB 4; Length 728;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QPGPVLSQP 10
|||||:
Db 444 QPGPILKRP 452

Search completed: February 16, 2003, 21:59:13
Job time : 3.70236 secs

XX Claim 1; Col 17-22; 12pp; English.
CC This cDNA sequence represents a human tata-binding protein associated
CC factor, hTAFII105, isolated from Daudi cell nuclear extracts. Tightly
CC associated subunits (TAF's) are components of the transcription factor
CC TFIID and are thought to mediate transcriptional activation. This encoded
CC protein may be produced recombinantly from transformed host cells or
CC purified from human cells. hTAFII105 specific binding agents such as
CC specific antibodies could be used for diagnosis (e.g. genetic
CC hybridisation screens for hTAFII105 transcripts), therapy (e.g. gene
CC therapy to modulate hTAFII105 gene expression) and in the
CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
CC cell specific activators or other transcriptional regulators).
SQ Sequence 801 AA;
Query Match 100.0%; Score 562; DB 19; Length 801;
Best Local Similarity 100.0%; Pred. No. 1.6e-54;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AAGPVLSPAGIPGTSSSKQLFSLFHVYQOPSGNEKQVTTISHSSTLTIOKCGQKMP 60
DB 443 AAGPVLSPAGIPGTSSSKQLFSLFHVYQOPSGNEKQVTTISHSSTLTIOKCGQKMP 502
OY 61 VNTIIPTSQFPASIIKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 110
DB 503 VNTIIPTSQFPASIIKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 552
RESULT 2
ABG09468
ID ABG09468 standard; Protein; 801 AA.
AC ABG09468;
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #9459.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS73655.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
PS
XX Claim 20; SEQ ID No 39827; 103pp; English.
CC
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 801 AA;
Query Match 100.0%; Score 562; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 1.6e-54;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AAGPVLSPAGIPGTSSSKQLFSLFHVYQOPSGNEKQVTTISHSSTLTIOKCGQKMP 60
DB 443 AAGPVLSPAGIPGTSSSKQLFSLFHVYQOPSGNEKQVTTISHSSTLTIOKCGQKMP 502
OY 61 VNTIIPTSQFPASIIKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 110
DB 503 VNTIIPTSQFPASIIKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 552
RESULT 3
AA57279
ID AA57279 standard; Protein; 852 AA.
AC AA57279;
XX
XX 06-JUN-2000 (first entry)
DT
XX
XX Transcription factor subunit TAFII105 polypeptide.
DE
XX
XX TATA box-binding protein associated factor II 105; TAFII105; cancer;
KM transcription factor; apoptosis; cytostatic; immunosuppressive;
XX antiinflammatory; virucide; antibacterial.
XX
XX Homo sapiens.
OS
XX
XX
FH
FH Key Location/Qualifiers
FT Misc-difference 802
FT /label= G1x
FT /note= "encoded by TGA"
FT Misc-difference 834
FT /label= G1x
FT /note= "encoded by TAA"
XX
XX WO200012699-A1.
PN
XX
XX 09-MAR-2000.
PD
XX
XX 25-AUG-1999; 99WO-IL00464.
PF
XX
XX 27-AUG-1998; 98IL-0125971.
PR
XX
XX (YEDA) YEDA RES & DEV CO LTD.
PA
XX
XX Dikstein R, Yamit-hezi A;
PI
XX
XX WPI; 2000-256640/22.
DR N-PSDB; AA290465.
XX
XX Polypeptide encoding TATA box binding protein associated factor II 105
PT useful for treating e.g. cancers and inducing apoptosis has a dominant

PT negative effect on the normal biological activity of the binding protein -

PS Claim 7: Fig 2: 48pp: English.

XX This represents a polypeptide comprising a (modified) fragment (I) of a TATA box-binding protein associated factor II 105 (TAFII105). A pharmaceutical composition comprising (I) or the polynucleotide or an inhibitor or antagonist of (I) is useful for treating cancers and inducing apoptosis in pathological cells. The composition is also useful for treating autoimmune diseases, inflammatory processes and viral or bacterial infections.

CC Sequence 852 AA;

QY Query Match 100.0%; Score 562; DB 21; Length 852; Best local Similarity 100.0%; Pred. No. 1.8e-54; Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 443 ACPGCVLSQPGAGIPTGSSSKQLFSLFHVVOQPSGNGEKVTTISHSSTLTIOKCGQKTM 60

QY 61 VNTIIPTSOPRPASIIKQITLPGKIIISLQASPTQKNRIKENTSCFRDE 110

Db 503 VNTIIPTSQPPASIIKQITLPGKIIISLQASPTQKNRIKENTSCFRDE 552

RESULT 4

AA056494

ID AA056494 standard; Protein: 737 AA.

AC AA056494;

XX

DT 23-MAR-1995 (first entry)

XX

DE TATA-binding protein-associated factor hTAFII130.

XX

KW TATA-binding protein associated factor; hTAFII130; screening; diagnostic; therapeutic; gene transcription regulation.

XX

OS Homo sapiens.

XX

PN WO9417087-A.

XX

PD 04-AUG-1994.

XX

PF 28-JAN-1994; 94NO-US01114.

XX

PR 28-JUN-1993; 93US-0013412.

XX

PR 30-JUN-1993; 93US-0087119.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N; Tjian R, Wang E, Weinzierl ROJ;

XX

DR WPI: 1994-264019/32.

XX

DR N-PSDB; AA070731.

XX

PT TATA-binding protein associated protein factors - and corresponding nucleotide sequence and deriv. antibodies, useful in screening, diagnostics and therapeutics

XX

PS Disclosure: Page 142; 180pp: English.

XX

CC The TATA-binding protein associated factor hTAFII130 (including specific antibodies and fusion products) are used in drug screening, diagnostics and therapeutics. They are used in the development of specific biochemical assays for screening compounds that agonise or antagonise selected transcription factors involved in regulating gene expression associated with human pathology.

CC

SQ Sequence 737 AA;

QY Query Match 16.5%; Score 92.5; DB 15; Length 737; Best local Similarity 27.3%; Pred. No. 0.18; Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

Db 370 QP-VLSLTPQTVGVCKGQPTPLVIOQPPKPGALIRPQVTLTQPMVALRQ----- 422

QY 2 QPGCVLS--QPGAGIPTGSSSKQLFSLFHVVOQPSGNGEKVTTISHSSTLTIOKCGQKTM 59

Db 423 PNNRIMLTTPQQLQNLNPLQPVVPAVLPRGTALSAVSAQAAAAQKNKLEPGGSGFRD 482

QY 60 PVNTIIP-----SQPPASIIKQITLPGKIIISLQASPTQKNRIKENTSCFRD 109

Db 423 PNNRIMLTTPQQLQNLNPLQPVVPAVLPRGTALSAVSAQAAAAQKNKLEPGGSGFRD 482

QY 110 E 110

Db 483 D 483

RESULT 5

AA06084

ID AA06084 standard; Protein: 737 AA.

AC AA06084;

XX

DT 27-JAN-1997 (first entry)

XX

DE Human TATA-binding protein associated factor hTAFII130 protein.

XX

KW Human: TATA-binding protein; TBP associated factor; TFIID; holoenzyme; RNA polymerase II; transcription; messenger RNA; nuclear fraction; lambda-gli1; expression library.

XX

OS Homo sapiens.

XX

PN US5534410-A.

XX

PD 09-JUL-1996.

XX

PF 28-JAN-1993; 93US-0013412.

XX

PR 28-JAN-1994; 94US-0188582.

XX

PR 30-JUN-1993; 93US-0013412.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N; Tjian R, Wang E, Weinzierl ROJ;

XX

DR WPI: 1996-333245/33.

XX

DR N-PSDB; AAT42217.

XX

PT Screen for cpds. that bind human TATA-binding protein associated factor - by testing ability to bind to polypeptide fragments of the factor, useful as (ant)agonists of transcription factors involved in disease.

XX

PS Examples: Column 105-112; 86pp: English.

XX

CC This is the amino acid sequence of the human TATA-binding protein (TBP) associated factor (TAF) designated TAFII130. The protein is a component of the TFIID fraction required for reconstituting RNA polymerase II in vitro transcription activity. The encoded protein has an estimated mol. wt. of 130 kD by SDS-PAGE.

CC

CC The invention relates to purified proteins involved in transcription by RNA polymerase II, the RNA polymerase which transcribes messenger RNA. RNA polymerase II transcription proceeds in vitro upon addition of several nuclear fractions designated TFIID, B, D, E, F, H, I and J to RNA polymerase II holoenzyme. Fraction TFIID has been shown to contain a TBP and other TAFs. Purification of TFIID and separation of its components reveals 7 proteins ranging in size from 30-250 kD.

CC

CC Serum raised against the TFIID fraction allowed cloning of the corresp.

CC genes from lambda-gt11 expression libraries.
 XX
 SQ Sequence 737 AA;
 Query Match 16.5%; Score 92.5; DB 17; Length 737;
 Best Local Similarity 27.3%; Pred. No. 0.18;
 Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;
 OY 2 QGPPVLS--QPGACIPMGSSSKQLFSLFHYVQOPSGGNGKQVTTISSTLTTCCKQKTM 59
 DB 370 QP-PVLSLQPTQVGVGKGGQPPPLVIOQPPKPGALIRPPQVTLTQTPVALRQ----- 422
 OY 60 PVTITPT-----SQFPASILKQITLPGNKIL--SLQASPTQKNRIKENVTSQCFRD 109
 DB 423 PNRRLMTTPQQTQLNPLQPPVYKPAVLPGTALSAVSAQAQAQKNKLEPGGGSFRD 482
 OY 110 E 110
 DB 483 D 483
 RESULT 6
 ID AAW25019 standard; Protein: 737 AA.
 XX AAW25019;
 AC AAW25019;
 DT 08-OCT-1997 (first entry)
 DE TATA-binding protein associated factor, hTAFII130.
 XX
 DE TATA-binding protein associated factor; TAF; nuclear protein;
 KM RNA polymerase transcription; TATA-binding protein; TBP;
 KM Initiation.
 OS Homo sapiens.
 XX
 PN US5637686-A.
 PD 10-JUN-1997.
 XX
 PF 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-0188582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 PR 09-MAY-1996; 96US-0646715.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Comat L, Dynalact BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 DR WPI: 1997-319113/29.
 DR N-PSDB: AAT79595.
 XX
 PT Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs
 PS
 XX Claim 1; Column 111-116; 86pp; English.
 CC AAW25018 represents TATA-binding protein associated factor (TAF)
 CC polypeptide, hTAFII100 (mol. weight 100kD). TAF peptides derived
 CC from hTAFII30 alpha, hTAFII30 beta, hTAFII40, hTAFII70, hTAFII100,
 CC hTAFII130, hTAFII250, hTAFII48 and hTAFII10 and nucleic acids encoding
 CC them, are used to modulate transcription, including transcription
 CC Initiation. TAFs are nuclear proteins involved in RNA polymerase I,
 CC II and III transcription. The peptides act by binding to a different
 CC TAF, an activator, or TBP (TATA-binding protein) or competitively
 CC inhibiting association of a TAF domain with another compound, typically
 CC a protein like TBP or another TAF, an activator, or DNA.

SQ Sequence 737 AA;
 Query Match 16.5%; Score 92.5; DB 18; Length 737;
 Best Local Similarity 27.3%; Pred. No. 0.18;
 Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;
 OY 2 QGPPVLS--QPGACIPMGSSSKQLFSLFHYVQOPSGGNGKQVTTISSTLTTCCKQKTM 59
 DB 370 QP-PVLSLQPTQVGVGKGGQPPPLVIOQPPKPGALIRPPQVTLTQTPVALRQ----- 422
 OY 60 PVTITPT-----SQFPASILKQITLPGNKIL--SLQASPTQKNRIKENVTSQCFRD 109
 DB 423 PNRRLMTTPQQTQLNPLQPPVYKPAVLPGTALSAVSAQAQAQKNKLEPGGGSFRD 482
 OY 110 E 110
 DB 483 D 483
 RESULT 7
 ID ABB63947
 XX ABB63947 standard; Protein: 3726 AA.
 XX ABB63947;
 AC ABB63947;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 18633.
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 18633.
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 DR WPI: 2001-656860/75.
 DR N-PSDB: ABL08050.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 18633; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 3726 AA;
 Query Match 14.1%; Score 79.5; DB 22; Length 3726;
 Best Local Similarity 18.2%; Pred. No. 51;
 Matches 30; Conservative 24; Mismatches 50; Indels 61; Gaps 5;

QY 3 PGVYL-----SOPAGIPFGSSSKOLFSLFHYVQOPSGGN-----EKQVTT 42
DB 5 PGVYMGIRLPAPSKPSETPPSTSSAFVWNPVIOAVSSSAVNTTQAPSLSSGASF 64
QY 43 ISHSSTLTID-----KCGQAKT-----MPVNT--IIPTS-----QFP--- 71
DB 65 VSGAGTTLTALISPPPEQSFASKTGSEFKITYSSGGCPVGTASLIRPLOSSGFALLQLPGR 124
QY 72 --PASILKQITLPGNKILSLQASPTQKNRIKENVTSCFDE 110
DB 125 PVPSSILQHYVA-----SIQMKRESQNPQDKDETNSIKRE 158
RESULT 10
ID AAB03192 standard; Protein; 524 AA.
AC AAB03192;
XX
XX 23-OCT-2000 (first entry)
DE Rat Elf-1 polymorphic variant #3.
XX
XX Elf-1; transcriptional regulatory factor; rat; polymorphic variant;
KW immunological disease; allergic disease.
XX
XX Rattus sp.
OS
FH Key Location/Qualifiers
FT Misc-difference 144..160
FT /note= "This region is apparently encoded by GGC"
XX
XX JP2000135088-A.
PN
XX 16-MAY-2000.
PD
XX 30-OCT-1998; 98JP-0309595.
PE
XX 30-OCT-1998; 98JP-0309595.
PR
XX (ASAK) ASAH I BREWERIES LTD.
PA
XX MPI; 2000-425802/37.
DR N-PSDB; AAA61647.
XX
XX A transcription regulating factor Elf-1 cDNA and its polymorphism -
PT
XX Claim 3; Page 17-18; 21pp; Japanese.
PS
XX
XX The invention relates to three polymorphic variants of a rat
CC transcriptional regulatory factor Elf-1 (AAB03190-H03192) and nucleic
CC acids encoding them (AAA61645-A61647). The invention also encompasses
CC the use of Elf-1 for regulating the transcription of a gene in an
CC expression vector. The cDNA encoding rat Elf-1 variant #1 (AAA61645)
CC was isolated using PCR primers (AAA61648-A61649) based on mouse Elf-1
CC cDNA, and the rat Elf-1 variant #2 and #3 cDNAs (AAA61646-A61647) were
CC subsequently isolated. The Elf-1 cDNA and protein sequences are
CC potentially useful for the prevention and treatment of immunological and
CC allergic diseases mediated by Elf-1. The present sequence represents
CC a variant (#3) of rat Elf-1.
XX
XX
SQ Sequence 524 AA.
Query Match 13.2%; Score 74; DB 21; Length 524;
Best Local Similarity 30.9%; Pred. No. 14;
Matches 30; Conservative 11; Mismatches 38; Indels 18; Gaps 5;

DB 324 GIRTQASTQVP-----VWSPGNOQLHTVTTLQTVP 354
RESULT 11
ID AAB03191 standard; Protein; 579 AA.
AC AAB03191;
XX
XX 23-OCT-2000 (first entry)
DE Rat Elf-1 polymorphic variant #2.
XX
XX Elf-1; transcriptional regulatory factor; rat; polymorphic variant;
KW immunological disease; allergic disease.
XX
XX Rattus sp.
OS
FH Key Location/Qualifiers
FT Misc-difference 133
FT /note= "Encoded by GAG"
FT Misc-difference 149
FT /note= "Encoded by GAR"
XX
XX JP2000135088-A.
PN
XX 16-MAY-2000.
PD
XX 30-OCT-1998; 98JP-0309595.
PE
XX 30-OCT-1998; 98JP-0309595.
PR
XX (ASAK) ASAH I BREWERIES LTD.
PA
XX MPI; 2000-425802/37.
DR N-PSDB; AAA61646.
XX
XX A transcription regulating factor Elf-1 cDNA and its polymorphism -
PT
XX Claim 2; Page 13-14; 21pp; Japanese.
PS
XX
XX The invention relates to three polymorphic variants of a rat
CC transcriptional regulatory factor Elf-1 (AAB03190-H03192) and nucleic
CC acids encoding them (AAA61645-A61647). The invention also encompasses
CC the use of Elf-1 for regulating the transcription of a gene in an
CC expression vector. The cDNA encoding rat Elf-1 variant #1 (AAA61645)
CC was isolated using PCR primers (AAA61648-A61649) based on mouse Elf-1
CC cDNA, and the rat Elf-1 variant #2 and #3 cDNAs (AAA61646-A61647) were
CC subsequently isolated. The Elf-1 cDNA and protein sequences are
CC potentially useful for the prevention and treatment of immunological and
CC allergic diseases mediated by Elf-1. The present sequence represents
CC a variant (#2) of rat Elf-1.
XX
XX
SQ Sequence 579 AA.
Query Match 13.2%; Score 74; DB 21; Length 579;
Best Local Similarity 30.9%; Pred. No. 16;
Matches 30; Conservative 11; Mismatches 38; Indels 18; Gaps 5;

QY 2 QPGPVLS--QPGAPITPGSSSKOLFSLFHYVQOPSGGNEKQVTTISSTLTITKCGCKTM 59
DB 326 QPSEVLRVQPSQAPYPT--QLFRVHVYQVPQVAIPPEATTIT--STWQERAGSSVQ 378
QY 60 PVNTIIPTSQFPASILKQITLPGNKIL--SLQASP 93
DB 379 GIRTQASTQVP-----VWSPGNOQLHTVTTLQTVP 409
RESULT 12
ID AAB03190 standard; Protein; 615 AA.
XX

AC AAB03190;
 XX 23-OCT-2000 (first entry)
 XX
 XX Rat E1f-1 polymorphic variant #1.
 DE
 XX E1f-1; transcriptional regulatory factor; rat; polymorphic variant;
 KW immunological disease; allergic disease.
 XX
 OS Rattus sp.
 XX
 PN JP2000135088-A.
 XX
 PD 16-MAY-2000.
 XX
 PE 30-OCT-1998; 98JP-0309595.
 XX
 PR 30-OCT-1998; 98JP-0309595.
 XX
 PA (ASAK) ASAMI BREWERIES LTD.
 XX
 DR WPI: 2000-425802/37.
 DR N-PSDB: AAA61645.
 XX
 PT A transcription regulating factor E1f-1 cDNA and its polymorphism -
 XX
 PS Claim 1; Page 8-10; 21pp; Japanese.
 XX
 CC The invention relates to three polymorphic variants of a rat
 CC transcriptional regulatory factor E1f-1 (AAB03190-B03192) and nucleic
 CC acids encoding them (AAA61645-A61647). The invention also encompasses
 CC the use of E1f-1 for regulating the transcription of a gene in an
 CC expression vector. The cDNA encoding rat E1f-1 variant #1 (AAA61645)
 CC was isolated using PCR primers (AAA61648-A61649) based on mouse E1f-1
 CC cDNA, and the rat E1f-1 variant #2 and #3 cDNAs (AAA61646-A61647) were
 CC subsequently isolated. The E1f-1 cDNA and protein sequences are
 CC potentially useful for the prevention and treatment of immunological and
 CC allergic diseases mediated by E1f-1. The present sequence represents
 CC a variant (#1) of rat E1f-1.
 XX
 SQ Sequence 615 AA;
 QY Query Match 13.2%; Score 74; DB 21; Length 615;
 DB Best Local Similarity 30.9%; Pred. No. 18;
 DB Matches 30; Conservative 11; Mismatches 38; Indels 18; Gaps 5;
 QY 2 QPGVLS--QPGAGPTGSSSKQLSLFHVVOQPSGNGEKQVTTISHSTLTIOKCGKTM 59
 DB 362 QPSEVLRTVQPSQAPVPT--QLFRVHVVPVQVAIPPEARIT---STMGEAAGSSVQ 414
 QY 60 PVNTIIPTSOPPPASILKQITLPGNKIL---SLQASP 93
 DB 415 GIRTQASTQV-----VVVSPGNQQLHTVLTQIVP 445
 RESULT 13
 ABB50171
 ID ABB50171 standard; Protein: 591 AA.
 AC ABB50171;
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE Human transcription factor TRFX-22.
 XX
 KW Human; transcription factor; TRFX; cell proliferative disease;
 KW autoimmune disease; inflammation; neurological disease;
 KW developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;
 KW neuroprotective; antiinflammatory; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200172777-A2.

XX 04-OCT-2001.
 PD
 XX 13-MAR-2001; 2001WO-US08117.
 PF
 XX 13-MAR-2000; 2000US-0188986.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Hillman JI, Baughn MR, Yue H, Lal P, Lu DAM, Patterson C;
 PI Azimzai Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;
 PI Reddy R;
 DR WPI: 2001-570896/64.
 DR N-PSDB: ABA82995.
 XX
 XX Novel transcription factor polypeptides, used to treat diseases
 PT associated with altered activity and expression of TRFX, and to screen
 PT for agents capable of modulating its activity -
 XX
 PS Claim 1; Pages 164-166; 327pp; English.
 XX
 CC The present sequence is the protein sequence for a human transcription
 CC factor. The transcription factor and its coding sequence are useful in
 CC the diagnosis, treatment and prevention of diseases associated with
 CC altered expression of the transcription factor e.g. cell proliferative,
 CC autoimmune/inflammatory, neurological and developmental disorders. A
 CC number of specific disorders/diseases are given in the specification,
 CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,
 CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
 CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,
 CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative
 CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's
 CC disease, stroke, and viral, bacterial, fungal and protozoal infections.
 XX
 SQ Sequence 591 AA;
 QY Query Match 13.0%; Score 73; DB 22; Length 591;
 DB Best Local Similarity 28.9%; Pred. No. 22;
 DB Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;
 QY 1 AOPGPVLS--OPAGIPTGSSSKQLSLFHVVOQ---PSGNGEKQVTTISHSTLTIOK 54
 DB 324 AQPSEVLRTVQPSQAPVPT--QLFRVHVVPVQVAIPPEARIT---EAKRTSTMDELINSSVQ 388
 QY 55 GOKTPVNTIIPTSOPPPASILKQITL-----PGNKILSLQASPTOK- 96
 DB 389 SIRTQAPTVQPVVSPNNQLHTVLTQIVPLTVIASTDPASAGTSGKFTLQAISSQP 448
 QY 97 -NRIRENV 103
 DB 449 MTVLKENV 456
 RESULT 14
 AAM38686
 ID AAM38686 standard; Protein: 595 AA.
 AC AAM38686;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 1831.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.

```

XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AA157842.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Example 3; SEQ ID NO 1831; 10078bp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AA157798-AA161369) with neurotropic,
XX CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Actin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 595 AA;
XX
XX Query Match 13.0%; Score 73; DB 22; Length 595;
XX Best Local Similarity 28.9%; Pred. No. 22;
XX Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;
XX
XX QY 1 ACPGVLAS--QPAIGPTGSSKQLFSLFHVQD---PSGNGEKQVTTISHSSTLTIOKC 54
XX DB 338 ACPSEVLRVQPTQSPYPT---QLFRVHVQPVQAVPEG--EAARTSTMQDETLLNSVQ 392
XX QY 55 GOKTMEVNTIIPTSOPPPASILKQITL-----PSNKILSLQASPTQK- 96
XX DB 393 SIRTIOAPTOVPVSPRNQQLHTVTLQTVPLTTVASTDPDSAGTSGKFFILQAISSQP 452
XX QY 97 -NRIKENV 103
XX DB 453 MIVLKENV 460
XX
XX RESULT 15
XX ID AA157842 standard; Protein; 619 AA.
XX AC AA157842;
XX XX
XX DT 22-OCT-2001 (first entry)

```

```

XX DE Human polypeptide SEQ ID NO 1830.
XX XX
XX KW Human: neurotropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW Leukemia.
XX OS Homo sapiens.
XX OS
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR N-PSDB; AA157841.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Example 3; SEQ ID NO 1830; 10078bp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AA157798-AA161369) with neurotropic,
XX CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Actin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 619 AA;
XX
XX Query Match 13.0%; Score 73; DB 22; Length 619;
XX Best Local Similarity 28.9%; Pred. No. 23;
XX Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;
XX
XX QY 1 ACPGVLAS--QPAIGPTGSSKQLFSLFHVQD---PSGNGEKQVTTISHSSTLTIOKC 54
XX DB 362 ACPSEVLRVQPTQSPYPT---QLFRVHVQPVQAVPEG--EAARTSTMQDETLLNSVQ 416
XX QY 55 GOKTMEVNTIIPTSOPPPASILKQITL-----PSNKILSLQASPTQK- 96
XX DB 417 SIRTIOAPTOVPVSPRNQQLHTVTLQTVPLTTVASTDPDSAGTSGKFFILQAISSQP 476
XX QY 97 -NRIKENV 103

```

Thu Feb 20 16:36:44 2003

us-09-763-909-2_copy_443_552.rag

Page 9

Db 477 MFTVLKENV 484

Search completed: February 16, 2003, 21:55:10
Job time : 10.4325 secs

This Page Blank (uspto)

Db 503 VNTIITSPQPPASILKQITLPGNKILSLQASPTQKRIKENVTSRDE 552

RESULT 2
US-08-188-582-16
; Sequence 16, Application US/08188582
; Patent No. 5534410

GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-188-582-16

Query Match 16.5%; Score 92.5; DB 1; Length 737;
Best Local Similarity 27.3%; Pred. No. 0.0046;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

OY 2 QPQPVLS--OPAGIPGSSSKQLFSLFHVQPSGNGEKQVTTISHSSTLTQKCGQKTM 59
Db 370 QP-PVLSLTQPTGVGKQGPPLVYIQPPKPGALIRPPVTLTOTPMVALRO----- 422

OY 60 PVTIITP-----SQPPASILKQITLPGNKIL--SLQASPTQKRIKENVTSRDE 109
Db 423 PNRIMLTTPQOIQLNLPDPVYKPAVLPGTALSAVSAQAAMAAKKNKLEKGGGSFRD 482

OY 110 E 110
Db 483 D 483

RESULT 3
US-08-646-715-16
; Sequence 16, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:

APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-16

Query Match 16.5%; Score 92.5; DB 1; Length 737;
Best Local Similarity 27.3%; Pred. No. 0.0046;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

OY 2 QPQPVLS--OPAGIPGSSSKQLFSLFHVQPSGNGEKQVTTISHSSTLTQKCGQKTM 59
Db 370 QP-PVLSLTQPTGVGKQGPPLVYIQPPKPGALIRPPVTLTOTPMVALRO----- 422

OY 60 PVTIITP-----SQPPASILKQITLPGNKIL--SLQASPTQKRIKENVTSRDE 109
Db 423 PNRIMLTTPQOIQLNLPDPVYKPAVLPGTALSAVSAQAAMAAKKNKLEKGGGSFRD 482

OY 110 E 110
Db 483 D 483

RESULT 4
US-09-134-001C-3773
; Sequence 3773, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C

;; CURRENT FILING DATE: 1998-08-13
;; PRIOR APPLICATION NUMBER: US 60/064,964
;; PRIOR FILING DATE: 1997-11-08
;; PRIOR APPLICATION NUMBER: US 60/055,779
;; PRIOR FILING DATE: 1997-08-14
;; NUMBER OF SEQ ID NOS: 5674
;; SEQ ID NO 3773
;; LENGTH: 412
;; TYPE: PRT
;; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3773

Query Match 12.98; Score 72.5; DB 4; Length 412;
Best Local Similarity 25.0%; Pred. No. 0.69;
Matches 28; Conservative 17; Mismatches 50; Indels 17; Gaps 4;

QY 6 VLSQPGIPGSSSKOLFSLFHVVOOPSGGNE--KQVTTIS-----HSTLTIOKCGOKTM 59
DB 170 VIAGAGGLPVGTGKTLMLSGGIDSPVAGLEVWKGVTEAIEHFHSPTSEKAKDKVI 229
QY 60 PVNTI-----IPTSGPPASILKQITLPKNKILSLQASPTOKRIKENYV 104
DB 230 ELTFLIAERVGPILKLHVPFTEIQKI-----NKVYHPRVTMTSTRMMMRIS 277

RESULT 5
US-09-514-247A-10
; Sequence 10, Application US/09514247A
; Patent No. 6365361
; GENERAL INFORMATION:
; APPLICANT: TANABE SEIYAKU CO. LTD.
; APPLICANT: TANIGUCHI, Tomoyasu
; APPLICANT: MIZUKAMI, Junko
; TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPA
; FILE REFERENCE: TANIGUCHI-6
; CURRENT APPLICATION NUMBER: US/09/514,247A
; CURRENT FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: PCT/JP98/03734
; PRIOR FILING DATE: 1998-08-24
; PRIOR APPLICATION NUMBER: JP231084/1997
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 10
; LENGTH: 2442
; TYPE: PRT
; ORGANISM: human
US-09-514-247A-10

Query Match 12.08; Score 67.5; DB 4; Length 2442;
Best Local Similarity 22.2%; Pred. No. 41;
Matches 30; Conservative 19; Mismatches 41; Indels 45; Gaps 5;

QY 2 QPGVYLSPPAIP-----TGSSSKOLFSLFHVVOOPSGGNE--KQV----- 40
DB 1910 QPSVSMSPAPFPVARTQPTVTSTGKPTQVPAAPPAVAARIEAEOOO 1969
QY 41 -----TTISHSS-----TLTIKCGOKTTPVNTIIPTSOPPPASILKO 78
DB 1970 OHLXRVNINNSMPPGRTGCTPGSGMAPVSLNVRPNQVSGPVMSMPGQMOQAPLPHQ 2029
QY 79 ITLPG--NKILSLQA 91
DB 2030 QPMGLPRPVISMQA 2044

RESULT 6
US-08-588-821-2
; Sequence 2, Application US/08588821
; Patent No. 5712097
; GENERAL INFORMATION:
; APPLICANT: Kern, Scott E.
; APPLICANT: Hahn, Stephan A.

;; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, DPC4
;; NUMBER OF SEQUENCES: 91
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson P.C.
;; STREET: 4225 Executive Square, Suite 1400
;; CITY: La Jolla
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,821
FILING DATE: 19-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/079001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 552 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-588-821-2

Query Match 11.74; Score 66; DB 1; Length 552;
Best Local Similarity 32.8%; Pred. No. 7.1;
Matches 22; Conservative 7; Mismatches 28; Indels 10; Gaps 3;

QY 8 SQPACIPGSSSKOLFSLFHVVOOPSGGNEK-----QVTTISHSSTLTIOKCGOKTTPVN 62
DB 223 SQPAILDGSHE---GLDQIASGPQPOQONGFTGQATYHNHSTTT--WTGSTRATYV 277
QY 63 TITPTSQ 69
DB 278 PNLPHHQ 284

RESULT 7
US-08-915-214-2
; Sequence 2, Application US/08915214
; Patent No. 5814457
; GENERAL INFORMATION:
; APPLICANT: Kern, Scott E.
; APPLICANT: Hahn, Stephan A.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, DPC4
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,214
FILING DATE: 20-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/588,821

APPLICANT: Kawabata, Masahiro
TITLE OF INVENTION: SMAD6 AND USES THEREOF
FILE REFERENCE: L0461/7038
CURRENT APPLICATION NUMBER: US/09/096,776B
CURRENT FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: US 60/049,990
PRIOR FILING DATE: 1997-06-13
PRIOR APPLICATION NUMBER: US 60/053,040
PRIOR FILING DATE: 1997-07-18
PRIOR APPLICATION NUMBER: US 60/066,173
PRIOR FILING DATE: 1997-11-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 552
TYPE: PRF
ORGANISM: Homo sapiens
US-09-096-776B-9

Query Match 11.7%; Score 66; DB 4; Length 552;
Best Local Similarity 32.8%; Pred. No. 7.1;
Matches 22; Conservative 7; Mismatches 28; Indels 10; Gaps 3;

QY 8 SQPAGIPTGSSKQLFLFHVVOOPSGGNEK-----QVTTISHSTLTLCCKGCKTAPVN 62
DB 223 SQPASILGGSHE---GLDQIASGPQPQOQNGFTGQPATYHNHSTTT--WTGSRITAPYT 277

QY 63 TIPTSQ 69
DB 278 PNLPHHQ 284

RESULT 11
US-09-312-748-4
Sequence 4; Application US/09312748
Patent No. 6423491
GENERAL INFORMATION:
APPLICANT: HOME, JAMES R.
APPLICANT: ALTONEN, LAURI A.
TITLE OF INVENTION: A GENE FOR JUVENILE POLYPOSIS
FILE REFERENCE: IOWA:021
CURRENT APPLICATION NUMBER: US/09/312,748
CURRENT FILING DATE: 1999-05-13
EARLIER APPLICATION NUMBER: 60/085,312
EARLIER FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 552
TYPE: PRF
ORGANISM: Homo sapiens
FEATURE:
US-09-312-748-4

Query Match 11.7%; Score 66; DB 4; Length 552;
Best Local Similarity 32.8%; Pred. No. 7.1;
Matches 22; Conservative 7; Mismatches 28; Indels 10; Gaps 3;

QY 8 SQPAGIPTGSSKQLFLFHVVOOPSGGNEK-----QVTTISHSTLTLCCKGCKTAPVN 62
DB 223 SQPASILGGSHE---GLDQIASGPQPQOQNGFTGQPATYHNHSTTT--WTGSRITAPYT 277

QY 63 TIPTSQ 69
DB 278 PNLPHHQ 284

RESULT 12
US-09-134-001C-4461
Sequence 4461; Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4461
LENGTH: 451
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4461

Query Match 11.7%; Score 65.5; DB 4; Length 451;
Best Local Similarity 24.6%; Pred. No. 6.1;
Matches 31; Conservative 20; Mismatches 40; Indels 35; Gaps 5;

QY 6 VLSOPAGIPTGSSKQLFLFHVVOOPSGGNEKQVTTISHSTL-----TIOKC 54
DB 75 VKSRATSIKRAKTSKQALS-----FKTSSTKTTTAKKSTTVAKTTTTRKQPTIRKS 127

QY 55 G-----OKTMEVN-----TIPTSQPPASILKQTLTPGNKILSLQASPTOKNRI 99
DB 128 STTSRSTKMTPTSVKRTTSKATATVPSKAKISRTKQQT--KSHSTSVAKNTTQLSKT 185

QY 100 KENVTS 105
DB 186 KSPSTS 191

RESULT 13
US-07-952-853-22
Sequence 22; Application US/07952853
Patent No. 5863783
GENERAL INFORMATION:
APPLICANT: Van Heuvel, Margaretha
APPLICANT: Bakhuys, Janna G.
APPLICANT: Coustel, Yves
APPLICANT: Harder, Abraham
APPLICANT: De Graaff, Leendert H.
APPLICANT: Filpphi, Michel J. A.
APPLICANT: Van Der Veen, Peter
APPLICANT: Visser, Jacob
APPLICANT: Andreoli, Peter M.
TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
TITLE OF INVENTION: MOLECULES
TITLE OF INVENTION: ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (ERO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/952,853
FILING DATE: 19921125
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 246152003500
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600

TELEFAX: 415-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 628 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-952-853-22

Query Match 11.7%; Score 65.5; DB 2; Length 628;
Best Local Similarity 26.2%; Pred. No. 9.9;
Matches 22; Conservative 13; Mismatches 30; Indels 19; Gaps 3;

Db 12 GIPGSSSKQLPSLFHVVOQPSG-----GNEKQVTTTSHSSTLTIOKCGQKTP 60
123 GIPVDGS--EFHSSFWIKDYSGDITVRLVGNVTGTEGSTTTHST-----ADNFTQ 174

QY 61 VNTIIPTSOPPASILKQITLPGN 84
175 ASVKFPTTKAPDGNVLYELTVDS 198

Db 175 ASVKFPTTKAPDGNVLYELTVDS 198

RESULT 14
US-08-914-848-22
Sequence 22, Application US/08914848
Patent No. 5989887
GENERAL INFORMATION:
APPLICANT: Van Heuvel, Margaretha
APPLICANT: Bakhuys, Janna G.
APPLICANT: Coutel, Yves
APPLICANT: Harder, Abraham
APPLICANT: De Graaff, Leendert H.
APPLICANT: Eliphil, Michel J. A.
APPLICANT: Van Der Veen, Peter
APPLICANT: Visser, Jacob
APPLICANT: Andreoli, Peter M.
TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DIGRADING ENZYMES OF FUNGAL
TITLE OF INVENTION: ORIGIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/914,848
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/952,853
FILING DATE: 25-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 246152003500
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600
TELEFAX: 415-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 628 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-914-848-22

Query Match 11.7%; Score 65.5; DB 2; Length 628;
Best Local Similarity 26.2%; Pred. No. 9.9;
Matches 22; Conservative 13; Mismatches 30; Indels 19; Gaps 3;

QY 12 GIPGSSSKQLPSLFHVVOQPSG-----GNEKQVTTTSHSSTLTIOKCGQKTP 60
123 GIPVDGS--EFHSSFWIKDYSGDITVRLVGNVTGTEGSTTTHST-----ADNFTQ 174

Db 123 GIPVDGS--EFHSSFWIKDYSGDITVRLVGNVTGTEGSTTTHST-----ADNFTQ 174

QY 61 VNTIIPTSOPPASILKQITLPGN 84
175 ASVKFPTTKAPDGNVLYELTVDS 198

Db 175 ASVKFPTTKAPDGNVLYELTVDS 198

RESULT 15
US-09-098-901-2
Sequence 2, Application US/09098901B
Patent No. 6218144
GENERAL INFORMATION:
APPLICANT: Scott, Mathew
APPLICANT: Sisson, John C.
TITLE OF INVENTION: Costal2 Genes and their Uses
FILE REFERENCE: SUN-65P
CURRENT APPLICATION NUMBER: US/09/098,901B
EARLIER FILING DATE: 1998-06-17
EARLIER APPLICATION NUMBER: 60/051,347
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1201
TYPE: PRT
ORGANISM: D. Melanogaster
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)-(1201)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-098-901-2

Query Match 11.7%; Score 65.5; DB 4; Length 1201;
Best Local Similarity 22.0%; Pred. No. 26;
Matches 26; Conservative 18; Mismatches 43; Indels 31; Gaps 3;

QY 6 VLSQPMGIPPGS-----SKQLPSLFHVVO-----PSGNEKQ----- 39
1081 LIRAPKPMPTGSLADRYKDEQSGRNIFAKFVLRYSAAAGSGSTAEESTALIDS 1140

Db 1081 LIRAPKPMPTGSLADRYKDEQSGRNIFAKFVLRYSAAAGSGSTAEESTALIDS 1140

QY 40 -----VTTTSHSSTLTIOKCGQKTPVNTIIPTSOPPASILKQITLPGKTIISLDS 92
1141 TTTATVTTTSTTTGAVGKVKALVSFRPEQLRLMPATATKVRORKNKTIIDDS 1198

Db 1141 TTTATVTTTSTTTGAVGKVKALVSFRPEQLRLMPATATKVRORKNKTIIDDS 1198

Search completed: February 16, 2003, 22:03:01
Job time : 4.82655 secs

;
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5708
LENGTH: 215
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5708

Query Match 12.2%; Score 68.5; DB 9; Length 215;
Best Local Similarity 25.2%; Pred. No. 3.2;
Matches 26; Conservative 19; Mismatches 35; Indels 23; Gaps 5;
DB 19 SKQSLFHVQPSGNGKQVTTSHSST-----LTIQKCGQRTMPVNTIIPTSOP- 70
68 TSHISLV-----QPLKGPPEAVSTIRATSSRYVERKHCHAMACESEST---GWLCPFAFE 120
QY 71 -----PPAS-----ILKQITLPGNKIISLQASPTQKNRIKENYVS 105
DB 121 VTNSPATSDSLAKARCAPDSRVASVAPSPKPTALSTVSA 163

RESULT 3
US-09-893-519A-14
Sequence 14, Application US/09893519A
Publication No. US20030027243A1
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BUDMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESTIVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/16548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:

NAME/KEY: misc_feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CAAT72189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)
US-09-893-519A-14

Query Match 12.1%; Score 68; DB 9; Length 1023;
Best Local Similarity 29.8%; Pred. No. 27;
Matches 28; Conservative 11; Mismatches 45; Indels 10; Gaps 4;
DB 5 PVLSPAGIPITGSSSKQSLFHVQPSGNGKQVTTSHSSTLTIQKCG-----QKTP 60
432 PRLPQPNPTNIGNQLPPGVLVRSNG-----QLMTIQQLAQMOQAHAQPTTMA 487
QY 61 VNTIIPTSOPPASITLQITLPGNKIISLQASPT 94
DB 488 PRATPTSA-PPVQT-STVQAPGPTIARQVTF 519

RESULT 4
US-10-109-886-10
Sequence 10, Application US/10109886
Patent No. US20020119499A1
GENERAL INFORMATION:
APPLICANT: TANABE SEIYAKU CO. LTD.
APPLICANT: TANIGUCHI, Tomoyasu
APPLICANT: MIZUKAMI, Junko
TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
FILE REFERENCE: TANIGUCHI-6
CURRENT APPLICATION NUMBER: US/10/109,886
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: 09/514,247
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: PCT/JP98/03734
PRIOR FILING DATE: 1998-08-24
PRIOR APPLICATION NUMBER: JP231084/1997
PRIOR FILING DATE: 1997-08-27
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 10
LENGTH: 2442
TYPE: PRT
ORGANISM: human
US-10-109-886-10

Query Match 12.0%; Score 67.5; DB 12; Length 2442;
Best Local Similarity 22.2%; Pred. No. 94;
Matches 30; Conservative 19; Mismatches 41; Indels 45; Gaps 5;
QY 2 QPGPVLSPAGIP-----TGSSSKQSLFHVQPSGNGNE--KQV----- 40
DB 1910 QPSVSMSPAGPSPVARQTPTTVSTGKPTSOVAPAPPPAPPAVEARQIERAQOQ 1969
QY 41 -----TTSHSS-----TLTIQKCGQRTMPVNTIIPTSOPPASILKQ 78
DB 1970 QHLRYVNNINSMPPRGTMGTGTPGSQMAPVSLNVRBNVSGPVPMPGMOQAAPLPOQ 2029
QY 79 ITLPG--NKILSLQA 91
DB 2030 QPAPGLRPVISMQA 2044

RESULT 5
US-09-815-242-10928
Sequence 10928, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.

```
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10928
LENGTH: 375
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10928
```

```
Query Match 11.9%; Score 67; DB 10; Length 375;
Best Local Similarity 27.8%; Pred. No. 9.6;
Matches 20; Conservative 19; Mismatches 29; Indels 4; Gaps 3;
```

```
OY 13 IPGSSSKQLFSLFHVQOPSGNGEKQVTTISH--SSTLIQKCGQKTMPTVNTIIPISQF 70
DB 148 LTSCKKQKQEMQLDTIEKREGRDYYVGVTVNGKSTL-INQIIQGVGVQDVITTSQF 206
```

```
OY 71 PPASILKQITLP 82
DB 207 -PGTTLKIEIP 217
```

```
RESULT 6
US-09-815-242-12792
Sequence 12792, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
```

```
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12792
LENGTH: 748
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12792
```

```
Query Match 11.9%; Score 67; DB 10; Length 748;
Best Local Similarity 25.2%; Pred. No. 23;
Matches 27; Conservative 23; Mismatches 35; Indels 22; Gaps 6;
```

```
OY 8 SOPAGIPGSSSKQLF----SLFHVQOPSGNGEKQVTTISHST--LTIQKCGQKTMPT 60
DB 225 SGP-----SRKKRVFAENSSNNIVNHQADQOOLTEQTHNSVESENTEIEAGEVT-N 277
OY 61 VNTIIPTSQPPASILKQITLPKNTILSLQASPTQKNTKENVTSCF 107
DB 278 VSIYV-----PPLTILNQ---PAKQKATSKAEVQRKGOVLENTYLKDF 316
```

```
RESULT 7
US-09-815-242-12327
Sequence 12327, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12327
LENGTH: 792
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12327
```

```
Query Match 11.9%; Score 67; DB 10; Length 792;
Best Local Similarity 25.2%; Pred. No. 25;
Matches 27; Conservative 23; Mismatches 35; Indels 22; Gaps 6;
```

```
OY 8 SOPAGIPGSSSKQLF----SLFHVQOPSGNGEKQVTTISHST--LTIQKCGQKTMPT 60
DB 269 SGP-----SRKKRVFAENSSNNIVNHQADQOOLTEQTHNSVESENTEIEAGEVT-N 321
```

```

Oy      61 VNTTITPSOPPASILKOTLTPGNLTLSQASPTOKNRRIKEVNTSCF 107
      1: : : 11 : : 1 : : : : : : : : : : : : : : : : : :
Db      322 VSYV-----PPLTLNQ---PAKQATSKAEVQRKGVLNTLAD 360

RESULT 8
US-09-923-922-9
: Sequence 9, Application US/099233922
: Patent No. US20020106732A1
: GENERAL INFORMATION:
: APPLICANT: Miyazono, Kohel
: TITLE OF INVENTION: SMAD6 AND USBS THEREOF
: FILE REFERENCE: 10461/7120
: CURRENT APPLICATION NUMBER: US/09/923,922
: CURRENT FILING DATE: 2001-08-07
: PRIOR APPLICATION NUMBER: 09/096,776
: PRIOR FILING DATE: 1998-06-12
: PRIOR APPLICATION NUMBER: US 60/049,990
: PRIOR FILING DATE: 1997-06-13
: PRIOR APPLICATION NUMBER: US 60/053,040
: PRIOR FILING DATE: 1997-07-18
: PRIOR APPLICATION NUMBER: US 60/066,173
: PRIOR FILING DATE: 1997-11-18
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 9
: LENGTH: 552
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-923-922-9

Query Match      11.7%; Score 66; DB 10; Length 552;
Best Local Similarity 32.8%; Pred. No. 20;
Matches 22; Conservative 7; Mismatches 28; Indels 10; Gaps 3;

Oy      8 SOPAGIPNGSSSKOLFSLFHHVYQDPGSGNEK-----QVTTTSHSSTLTITQKGGOKTMPVN 62
      1: : : 11 : : 1 : : : : : : : : : : : : : : : : : :
Db      223 SOPAIIIGGSHSE---GLQLQISGFPQGGQONGFTGQPATYTHHNSITT--WTGSRAPYT 277

Oy      63 TITPSQ 69
      1: : 1
Db      278 PMLPHNQ 284

RESULT 9
US-09-964-899-33
: Sequence 33, Application US/09964899
: Patent No. US20020174446A1
: GENERAL INFORMATION:
: APPLICANT: Cohen, Dalia et al.
: TITLE OF INVENTION: Identification of Genes Involved in
: FILE REFERENCE: 4-31612 A
: CURRENT APPLICATION NUMBER: US/09/964,899
: CURRENT FILING DATE: 2001-09-27
: PRIOR APPLICATION NUMBER: 60/236,893
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: 60/298,309
: PRIOR FILING DATE: 2001-06-14
: NUMBER OF SEQ ID NOS: 53
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 33
: LENGTH: 1249
: TYPE: PRT
: ORGANISM: Homo Sapien
US-09-964-899-33

Query Match      11.7%; Score 66; DB 9; Length 1249;
Best Local Similarity 21.3%; Pred. No. 58;
Matches 27; Conservative 22; Mismatches 54; Indels 24; Gaps 5;

Oy      1 AOPGCVLSQ---PACIPNGSSSKOLFSLFHHVYQDPGSGNEKOVTTTSHSSTLTITQKCGOK 57

```

[illegible]

NAME/KEY: UNSURE
LOCATION: (12)
NAME/KEY: UNSURE
LOCATION: (449)
US-09-729-674-176

Query Match

11.4%; Score 64; DB 10; Length 832;

Best Local Similarity 26.2%; Pred. No. 57;

Matches 32; Conservative 14; Mismatches 46; Indels 30; Gaps 5;

QY 2 OPGPV--SOPAGIPGSSSKOLFSLFHHVQOPSGNEKQVTTISHSSTLTIOKCGQKMP 60
DB 206 QNGTVALPSEAGLATASCPITVSSVAASQGLCVTNTRTSSVRKQ---LFACVKTSP 261
QY 61 VNTIT-----PTSGPPASITKQITLPGNKILISQASPTQKN 97
DB 262 PAVIVSSVSTCSSLPSSVSAPIITSCAPPT-FLPAS-TSQQLSSQKMFSAVPPTKE 319
QY 98 RI 99
DB 320 KV 321

QY 98 RI 99
DB 320 KV 321

RESULT 15

US-09-864-761-43246

Sequence 43246, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecm1ca-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 43246
LENGTH: 432
TYPE: PRT
ORGANISM: Homo sapiens

FEATURE: MAP TO AC005859.1

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99

OTHER INFORMATION: EST HUMAN HIT: AM867810.1, EVALUATE 5.00e-13

OTHER INFORMATION: SWISSPROT HIT: Q28888, EVALUATE 4.00e-14

US-09-864-761-43246

Query Match

11.3%; Score 63.5; DB 10; Length 432;

Best Local Similarity 24.0%; Pred. No. 28;

Matches 31; Conservative 14; Mismatches 31; Indels 53; Gaps 6;

QY 10 PAGIPGSSSKOLFSLFHHVQOPSGNEKQVTTISHSSTLTIO-----KCGQ 56
DB 59 PGGIPPTNTNLT-----TINHPDISPASFHRDLHLVEIDFRCNC 99
QY 57 KTMPV---NTIIPTSQPPASI-----LKQITLPGNKIL-----SLQASPTQKNR 99
DB 100 VPIPLSGKNMCMCKIRLQIKRPSFSGLTYSYLDGNGLEIFQGLPSPQLLSLEANNI 159
QY 100 -----KENVT 104
DB 160 FSIRKENLTF 168

Search completed: February 16, 2003, 22:02:37
Job time: 5.16702 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:03:10 ; Search time 92 Seconds
(without alignments)
8526.947 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558
Sequence: 1 gggaccctgtgaccaaagt.....cttattacttactat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/2/lna/5A_COMB.seq:*
3: /cgn2_6/ptodata/2/lna/5B_COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6A_COMB.seq:*
5: /cgn2_6/ptodata/2/lna/6B_COMB.seq:*
6: /cgn2_6/ptodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	2534	99.1	2556	1 US-08-725-012-1	Sequence 1, Appl
2	338.2	13.2	3603	1 US-08-188-582-15	Sequence 15, Appl
3	338.2	13.2	3603	1 US-08-646-715-15	Sequence 15, Appl
4	142.2	5.6	4615	1 US-08-188-582-1	Sequence 1, Appl
5	142.2	5.6	4615	1 US-08-646-715-1	Sequence 1, Appl
6	57.6	2.3	7218	1 US-08-232-463-14	Sequence 14, Appl
7	41	1.6	2169	1 US-09-434-408-3	Sequence 3, Appl
8	37	1.4	2223	1 US-08-257-073-4	Sequence 4, Appl
9	36	1.4	998	1 US-09-122-4008-5	Sequence 5, Appl
10	35.4	1.4	397	3 US-09-253-691-3	Sequence 3, Appl
11	35.4	1.4	5163	3 US-08-700-651-1	Sequence 1, Appl
12	35.4	1.4	5163	3 US-08-928-361B-4	Sequence 4, Appl
13	35.4	1.4	5318	3 US-08-700-651-2	Sequence 2, Appl
14	35.4	1.4	5318	3 US-08-928-361B-3	Sequence 3, Appl
15	35.2	1.4	2030	4 US-09-512-342-1	Sequence 1, Appl
16	35	1.4	543	6 5273901-6	Patent No. 5273901
17	35	1.4	2190	4 US-09-625-188-19	Sequence 19, Appl
18	34.8	1.4	19307	3 US-08-836-022A-10	Sequence 10, Appl
19	34.8	1.4	19307	3 US-09-427-048A-10	Sequence 10, Appl
20	34.6	1.4	1778	1 US-08-416-870C-3	Sequence 3, Appl
21	34.4	1.3	387	4 US-09-615-192A-253	Sequence 253, App
22	34.4	1.3	1916	4 US-09-484-970B-6	Sequence 6, Appl
23	34.4	1.3	21338	4 US-08-961-527-20	Sequence 20, Appl
24	34.2	1.3	376	2 US-08-623-906A-18	Sequence 18, Appl
25	34.2	1.3	1008	4 US-09-527-345-1	Sequence 1, Appl
26	34.2	1.3	1170	4 US-08-973-005A-9	Sequence 9, Appl
27	34.2	1.3	2338	4 US-08-845-258-13	Sequence 13, Appl

C 28	34.2	1.3	2338	4 US-08-990-571-13	Sequence 13, Appl
C 29	34.2	1.3	2338	4 US-08-723-142A-13	Sequence 13, Appl
C 30	34.2	1.3	2338	4 US-09-528-784A-13	Sequence 13, Appl
31	34	1.3	3666	2 US-08-682-517-13	Sequence 2, Appl
32	34	1.3	3666	2 US-08-682-517-14	Sequence 14, Appl
33	34	1.3	4197	2 US-08-682-517-7	Sequence 7, Appl
34	34	1.3	4197	2 US-08-682-517-8	Sequence 8, Appl
35	34	1.3	7705	2 US-08-687-080-115	Sequence 115, Appl
36	33.8	1.3	1695	1 US-08-361-920-20	Sequence 20, Appl
37	33.8	1.3	1695	1 US-08-479-939-20	Sequence 20, Appl
38	33.8	1.3	1695	1 US-08-483-432-20	Sequence 20, Appl
39	33.8	1.3	2088	4 US-09-351-414-3	Sequence 3, Appl
C 40	33.2	1.3	9884	4 US-09-791-211-10	Sequence 8, Appl
41	33	1.3	21040	4 US-08-961-527-55	Sequence 55, Appl
42	32.8	1.3	1545	4 US-08-900-117A-2	Sequence 2, Appl
43	32.8	1.3	2088	1 US-08-331-394-1	Sequence 1, Appl
44	32.8	1.3	2088	1 US-08-250-858-1	Sequence 1, Appl
45	32.8	1.3	2088	1 US-08-446-915-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-08-725-012-1
: Sequence 1, Application US/08725012
: Patent No. 5710025
:
: GENERAL INFORMATION:
: APPLICANT: Dikstein, Rivka
: APPLICANT: Tjian, Robert
: TITLE OF INVENTION: B-Cell Specific Transcription Factor
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/725,012
: FILING DATE:
:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: B97-005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2556 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..2403
:
: US-08-725-012-1
:
: Query Match 99.1%; Score 2534; DB 1; Length 2556;
: Best Local Similarity 99.9%; Pred. No. 0;
: Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
```

Dh 1 GGGACCCCTGGTGACAAAGTGGCTCCGGTCAGGCGCCCTCTAAAGTACAGAGCGGCCCT 60
QY 61 AGGCTGGCTCTCTCTCAGATAGTCGCCGTGAAGGCCCCCAACACACAGACATCTCACTTT 120
Dh 61 AGGCTGGCTCTCTCTCAGATAGTCGCCGTGAAGGCCCCCAACACACAGACATCTCACTTT 120
QY 121 CTTGCTAATTTGGAGCTTCTCCAGAAACGGTTTGTATTTAAAGTAAACAGTGGCCGTTG 180
Dh 121 CTTGCTAATTTGGAGCTTCTCCAGAAACGGTTTGTATTTAAAGTAAACAGTGGCCGTTG 180
QY 181 ATGTTGGTATCTCCTCAGCAAACTGTAAACAAGCCGAGACACACAACTAACTAACCTCA 240
Dh 181 ATGTTGGTATCTCCTCAGCAAACTGTAAACAAGCCGAGACACACAACTAACTAACCTCA 240
QY 241 AGGCCAGCACTACGAGGAATCTCTCAAAAGCTAAAACTGTACAGTGGCCGAACTTACG 300
Dh 241 AGGCCAGCACTACGAGGAATCTCTCAAAAGCTAAAACTGTACAGTGGCCGAACTTACG 300
QY 301 TCACAAATTAATCAAGAAAGTGGCAGTACACACTGTTAAAAAATTTGGCAAAATAGSAACT 360
Dh 301 TCACAAATTAATCAAGAAAGTGGCAGTACACACTGTTAAAAAATTTGGCAAAATAGSAACT 360
QY 361 AGTGTGTAAACAAGTGTCCGAAAGCCCTTCTCAGTACAACTGTGGCTGTGCCAACAGT 420
Dh 361 AGTGTGTAAACAAGTGTCCGAAAGCCCTTCTCAGTACAACTGTGGCTGTGCCAACAGT 420
QY 421 RGTGTCACAGTTCACGTCCTGGAAGCCATGTGAATACGTACCTTACCTGAAAGCTTAAAG 480
Dh 421 -GTGCTACAGTTCACGTCCTGGAAGCCATGTGAATACGTACCTTACCTGAAAGCTTAAAG 480
QY 481 TTTGGGAGCATCATCCACTCTCTCAATTAAGCCCAATCTTAAAGCAGAGAACTCAGCAGC 540
Dh 481 TTTGGGAGCATCATCCACTCTCTCAATTAAGCCCAATCTTAAAGCAGAGAACTCAGCAGC 540
QY 541 TGTTCAGATTAATCTTTTTCGACAAATGTGAAGAAATGTGAAGAAATGTGAAGAAATGT 600
Dh 541 TGTTCAGATTAATCTTTTTCGACAAATGTGAAGAAATGTGAAGAAATGTGAAGAAATGT 600
QY 601 TGGAAATGTTAAATAAATACATGTAGTGTATGATACAGTCCCTCGAAATGGGGCAAAATGT 660
Dh 601 TGGAAATGTTAAATAAATACATGTAGTGTATGATACAGTCCCTCGAAATGGGGCAAAATGT 660
QY 661 GAAGAAGCTGTGGAAACAATTTTGGATGCAAAAATCGAAGCAGAAAGATTTTACTAGAA 720
Dh 661 GAAGAAGCTGTGGAAACAATTTTGGATGCAAAAATCGAAGCAGAAAGATTTTACTAGAA 720
QY 721 ACTGTATGTTGAACCTAAGCTTCCACCTCAGCTCAGCTGTTCTTCTTAAAGAAAG 780
Dh 721 ACTGTATGTTGAACCTAAGCTTCCACCTCAGCTCAGCTGTTCTTCTTAAAGAAAG 780
QY 781 CGTGTGTGCTTACGAACTTCTGCTTAATCCGAGAGTTCATCCAGCAATGTGTCTCA 840
Dh 781 CGTGTGTGCTTACGAACTTCTGCTTAATCCGAGAGTTCATCCAGCAATGTGTCTCA 840
QY 841 GCAGACTTCTAGTACATGCTATGCTACCTGCTACTACAAAGTAAACAATCTCTCTGT 900
Dh 841 GCAGACTTCTAGTACATGCTATGCTACCTGCTACTACAAAGTAAACAATCTCTCTGT 900
QY 901 GGTGACAACTACAGTGTCTCAAGCCAGTGTGAAGAGTCAATTAATGTTTCTGAGACAC 960
Dh 901 GGTGACAACTACAGTGTCTCAAGCCAGTGTGAAGAGTCAATTAATGTTTCTGAGACAC 960
QY 961 AGCACCAGAACTGTGCTGCAAGCTTGAACCCCACTGGTGTCCACTGGAGAGAA 1020
Dh 961 AGCACCAGAACTGTGCTGCAAGCTTGAACCCCACTGGTGTCCACTGGAGAGAA 1020
QY 1021 AGCTGAGTGTGACACTTCTATCTGTGGCCCACTGTGCAACAGAGAGAAACAAGC 1080
Dh 1021 AGCTGAGTGTGACACTTCTATCTGTGGCCCACTGTGCAACAGAGAGAAACAAGC 1080
QY 1081 TGGAACTGTTTGTCTCAGACTTCAAAACCACTTGTGACATCTGTGGCAACACAGTGAC 1140
Dh 1081 TGGAACTGTTTGTCTCAGACTTCAAAACCACTTGTGACATCTGTGGCAACACAGTGAC 1140

QY 1141 CACGCTCTACCTGCAACCTGAAAAGCCAGTTGTCTTGGAACACAGTAACTGTCCCT 1200
Dh 1141 CACGCTCTACCTGCAACCTGAAAAGCCAGTTGTCTTGGAACACAGTAACTGTCCCT 1200
QY 1201 TCCAGCAGTAACTTTTGGAGAACTTCAAGTGGAGCTATTGTGCTCATCTGTGAAGC 1260
Dh 1201 TCCAGCAGTAACTTTTGGAGAACTTCAAGTGGAGCTATTGTGCTCATCTGTGAAGC 1260
QY 1261 TGTGTTTCTTCTGCTGGAGCACAATCTGCAACCTGTTAATGAGACTCCAGTTCAAAAT 1320
Dh 1261 TGTGTTTCTTCTGCTGGAGCACAATCTGCAACCTGTTAATGAGACTCCAGTTCAAAAT 1320
QY 1321 CAAACTTGGCCAGCCGGGCCCTGTCTTTCACAAACAGCTGGAGTTCACAGGCACTTC 1380
Dh 1321 CAAACTTGGCCAGCCGGGCCCTGTCTTTCACAAACAGCTGGAGTTCACAGGCACTTC 1380
QY 1381 AAGCAGCACTATTTCAATGTTTCAAGTATGATGACAGCTTCAGAGCAATGAAA 1440
Dh 1381 AAGCAGCACTATTTCAATGTTTCAAGTATGATGACAGCTTCAGAGCAATGAAA 1440
QY 1441 ACAAGTACCAACAATTTCAATCTCAGCAATGACCAATTCAGAAATGTGGACAGAAAGC 1500
Dh 1441 ACAAGTACCAACAATTTCAATCTCAGCAATGACCAATTCAGAAATGTGGACAGAAAGC 1500
QY 1501 GATGCCAGTGAACCACTAATACCTACTAGTACAGTTCCTCAGCTTCATTTAAAGCA 1560
Dh 1501 GATGCCAGTGAACCACTAATACCTACTAGTACAGTTCCTCAGCTTCATTTAAAGCA 1560
QY 1561 AATTACCTGCTGCTGGAATTAATTTCTGCTACATGCAAGATCTCCACAGCAAGAAATA 1620
Dh 1561 AATTACCTGCTGCTGGAATTAATTTCTGCTACATGCAAGATCTCCACAGCAAGAAATA 1620
QY 1621 GAATTAAGAAAGAAATGTAAATCATCTTCCAGATGAGATGACATCAATGATGTGACTT 1680
Dh 1621 GAATTAAGAAAGAAATGTAAATCATCTTCCAGATGAGATGACATCAATGATGTGACTT 1680
QY 1681 CTATGCGAGGGGTACACCTTAATGAGAAATGCTCTCATCTTACAAACAACCTGTAAT 1740
Dh 1681 CTATGCGAGGGGTACACCTTAATGAGAAATGCTCTCATCTTACAAACAACCTGTAAT 1740
QY 1741 TGGTTGGCACACTCATTCAGTCAATGTAAGTAAAGCAATTTCTTTTATGGAGCTCTAC 1800
Dh 1741 TGGTTGGCACACTCATTCAGTCAATGTAAGTAAAGCAATTTCTTTTATGGAGCTCTAC 1800
QY 1799 TGGTTGGCACACTCATTCAGTCAATGTAAGTAAAGCAATTTCTTTTATGGAGCTCTAC 1958
Dh 1799 TGGTTGGCACACTCATTCAGTCAATGTAAGTAAAGCAATTTCTTTTATGGAGCTCTAC 1958
QY 1801 AAAAGAGAAATCTTACATTTGTAAAAAGCATGACATTAAGCAATTAATCTGTATGCTG 1860
Dh 1801 AAAAGAGAAATCTTACATTTGTAAAAAGCATGACATTAAGCAATTAATCTGTATGCTG 1860
QY 1861 TGAACCTGATCTCCCAAGCAACAGAGACGACTACAGGCTTCTTGAAGAAACCTGACTG 1920
Dh 1861 TGAACCTGATCTCCCAAGCAACAGAGACGACTACAGGCTTCTTGAAGAAACCTGACTG 1920
QY 1921 CAATTTGCTACAGTCAATGATGATCTTACAAGGCAATGAAATTAATCTGTATGCTG 1980
Dh 1921 CAATTTGCTACAGTCAATGATGATCTTACAAGGCAATGAAATTAATCTGTATGCTG 1980
QY 1919 CAATTTGCTACAGTCAATGATGATCTTACAAGGCAATGAAATTAATCTGTATGCTG 1978
Dh 1919 CAATTTGCTACAGTCAATGATGATCTTACAAGGCAATGAAATTAATCTGTATGCTG 1978
QY 1981 ATACCAAGTACAGCTCAAAATTTCTTGAAGAAAGCTGATCAATTTGAGCAACAGAGAAAG 2040
Dh 1981 ATACCAAGTACAGCTCAAAATTTCTTGAAGAAAGCTGATCAATTTGAGCAACAGAGAAAG 2040
QY 2041 AATTGGAAGAAAGAAATTTTACTTTAAGCAGCAAGAGTCTTCTTAATTAAGAGATC 2100
Dh 2041 AATTGGAAGAAAGAAATTTTACTTTAAGCAGCAAGAGTCTTCTTAATTAAGAGATC 2100
QY 2101 CAAACAGCTGAGATTTAAACAGAAAGCAAAGAGTTACAGCAATTTGAACTTGGACAGA 2160
Dh 2101 CAAACAGCTGAGATTTAAACAGAAAGCAAAGAGTTACAGCAATTTGAACTTGGACAGA 2160
QY 2099 CAAACAGCTGAGATTTAAACAGAAAGCAAAGAGTTACAGCAATTTGAACTTGGACAGA 2158
Dh 2099 CAAACAGCTGAGATTTAAACAGAAAGCAAAGAGTTACAGCAATTTGAACTTGGACAGA 2158
QY 2161 TACAGCATAGAGAGCTAATCTCAGAGCTTTCAGCTATTTGAGCAACAGAGAAAGAGAC 2220
Dh 2161 TACAGCATAGAGAGCTAATCTCAGAGCTTTCAGCTATTTGAGCAACAGAGAAAGAGAC 2220

APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Wang, Edith
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/646,715
FILING DATE: 09-MAY-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/188,582
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3603 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2214
US-08-646-715-15

Query Match 13.2%; Score 338.2; DB 1; Length 3603;
Best Local Similarity 63.9%; Pred. No. 6.3e-92;
Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

1586 TCCTGCTCACTTCAAGCATCTCTCTCTCAAGAAATAGAAATAAAGAGATGTAAACATCATG 1645
1377 TGCTCGGCAAGCAAGCATCTCTCTCTCAAGAAATAACTCAAGAGAGCTGGGGAGGTTG 1436
1646 CTTCGAGATGAGATGACATCATGATGACTCTATGAGGAGGGGTCAACCTTATGA 1705
1437 GTTTCGGAGAGATGACATCATGATGACTCTATGAGGAGGGGTCAACCTTATGA 1496
1706 AGAAATGCTGATCTTACCAACAACTCTGAATGGTGGACACTTATGAGTCATG 1765
1497 AGAAGTGCAGAAATATATGACCAAGAACTGTAATGGTGGACAGCTTAACGGGCTCTG 1556
1766 TAAAGATGAACATTTCTTTTATTGGAGCTTACAAAGAGAACTTACATTTGTAA 1825
1557 TAAAGATGAACATTTCTTTTATTGGAGCTTACAAAGAGAACTTACATTTGTAA 1616
1826 AAGCATGACATTAAGAAATTAAGTCTGATGCTGAATGATCTCCCAACACACACA 1885
1617 AAAACATGATTAAGAAATTAAGTCTGATGCTGAATGATCTCCCAACACACACA 1676
1886 GGAACGACTACAGGCTCTTACAAAAGAACTGACTGCAATTTGTCAGCATGCAATGACTAC 1945

1677 ACAAAGCTACAGATCTTTAGAGAAATATCAAGAAACAGCTACAGCAAGAACTTTTC 1736
1946 TTACAGGCAAGTGAATAATATACATCTGTGTAGTATACAGCTACAGCTCAATTTCT 2005
1737 TTACAGGATGACAGATATATGAGAGGCGAGTGTGCGGCGACAGCTCAAGTTTCT 1796
2006 TGAAGAGTGTATCAATTTGGAGAAAGCAAGAAAGATTTGGAGAAAGAAATGTTACT 2065
1797 TGAACAGCTTGTATCAATTTGGAGAAAGCAAGAAAGATTTGGAGAAAGAAATGTTACT 1856
2066 TAAAGCAAGCAAGTGTGTTCTATTAAGAAATCAAGCAAGCAAGTGTATTAAGCAAGAA 2125
1857 GAGGCGCAAGATCTGTGTGTAAGCAAGAAATCAAGCAAGTGTATTAAGCAAGAA 1916
2126 AGCCAAAGAGTTACAGCAATTTGAATCTTGCACAGATACAGATAGAGACCTTAATCTAC 2185
1917 GCGAAAGAGATGCAAGCAAGCAAGTGTGCGGCGACCAATATGACAGGCGGACCCCACTAC 1976
2186 AGCTCTGACATTTTGGACCAAGAAAGAAAGACCACTTGA-----ATC 2230
1977 AGCACTAGCAGCGATCGGCGCCAGAAAGAAAGAGTGTACTGTCCGGGCGGCGCTC 2036
2231 TGAATGAGGCGTTTAAAGCAACCTTCTTCTGGGACATCCAGCTGACAGCCAC 2290
2037 AGGAGCAGAGGCGTCCGCGCCCGCTCAGTGTCTCCAGCACTCGGCTGCGAACCC 2096
2291 CAACAGTTGATCTTCAAGAAATCAGAGATCTGCTCAGAGGAGCTTGAATTTGTAT 2350
2097 CAGACAGTTCAAGCAAGAAATCAGAGGCTCAACCTCAGAGGAGCTCAATTTTGT 2156
2351 GGAACAGAAAGGAGATGAATTTCTGAGCTCTTACTGCGCTTCTGAAGTAC 2409
2157 AGAAATGAACGTTGAGCAAGCAAGCTTCACTGCTCTCAAGAAAGCAATTTCTTAAGTAC 2215

RESULT 4
US-08-188-582-1
Sequence 1, Application US/08188582
Patent No. 5534410
GENERAL INFORMATION:
APPLICANT: Tjian, Robert
APPLICANT: Comai, Lucio
APPLICANT: Dynlacht, Brian D.
APPLICANT: Hoey, Timothy
APPLICANT: Ruppert, Siegfried
APPLICANT: Tanese, Naoko
APPLICANT: Weinzierl, Robert O.J.
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989

QY 1893 CTACGAGGCGCTTCAGAAAACTGACGCAATTGCTCAGCATGCAATGACTACTTACAG 1952
DB 2785 CTAGAAACACATCGTTGGAAGTTGGCTGTATATGCGGAGCACCGCATTTGATCATCAAG 2844
QY 1953 GCAAGTGAATAATTCATCTGTGTAGTATCAGAGTACAGCTCAATTTCTTGAAGAAG 2012
DB 2845 TTGATTCACAGCTATGAGCCCGCAAGAGATGTGGCGGTGAGATCAAGTTCTTGAGAG 2904
QY 2013 CTGATCAATGGAGAAAGCAAGAAAGATTTGGAGAAAGAAATGTTACTTAAAGCA 2072
DB 2905 CTGGAGAAAGGCGAGCAAGCAACAGAGAACTGGAGGTGATGCTGCTGGGCA 2964
QY 2073 GCCAAGAGTCTTCTAATAAAGATCCAGACGCTGAGATTAAGAGAAAGCCAA 2132
DB 2965 GCCAAGTACAGGTCGAGAGGTGGAAGATCCGAGACGCGCAAGATGAAGGAGGCGCAAG 3024
QY 2133 GAGTACAGCAATTTGCAATTTGACAGATACAGATAGAGAGCTTAATTCACAGCTCTT 2192
DB 3025 GAGATGCAACGCGCGAAATGAGAGATGTCGTCACAGATGCAATTCGAGCGGCTG 3084
QY 2193 GCAAGTATTTGACCAAGAAAGAGACCACTAGATCTGGAATTTGAGGCTTAAAGAC 2252
DB 3085 CAGGCAATTTGACCTGGAAAAAGCTGAACTGGACGCGCAACAGTCACTTGGGAGCG 3144
QY 2253 AACCTTCTTCTCTGCG---GACATCCAGGCTGACAGCACAACAGCTTGCATGCTCA 2309
DB 3145 GGTTCAGATGCGCGCGAGTGTAGCAGCTCGGATCTCGCGGACGACGCTTAAGCGCT 3204
QY 2310 AGAATCAGCAATCTGCTCAGGACTGTATATTTGTATGAAACAGAAAGGAGATG 2369
DB 3205 CGCATTAACGTGTGAACCTGCGGACATGCTCTTACATGGAAGCAAGGCGGATTC 3264
QY 2370 AAGTATTTCTGACTCTATACCTGGCCCTTGAAGTGACCACTCCACTTTCCATCCA 2428
DB 3265 TGTGCGAGTTGCATGCTGTTCAGACATACCTCAAGTATGCTGCTGCTGCCATCAA 3323

RESULT 6
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300
TELEFAX: (703)863-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PL-F15
US-08-232-463-14

Query Match 2.3%; Score 57.6; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 9.2e-07;
Matches 15; Conservative 226; Mismatches 155; Indels 0; Gaps 0;

QY 1885 AGCAAGCACTACGAGGCGCTCTAGAAAAAGCATGCAATTTGCTCAGCATGCAATGACTA 1944
DB 1436 ACRR 1377
QY 1945 CTACAAAGCAAGTGAATAATTCATCTCTGTAGTATACAGTCAACGCTCAATTTTC 2004
DB 1376 RRR 1317
QY 2005 TTGAAAAGCTGCATCAATTTGACAGACAGCAAGAAAGATTGGAAGAAAGAAATTTAC 2064
DB 1316 RRR 1257
QY 2065 TTAAGCAGCAGCAGAGCGTTCTTAATAAGAGATCCAGAAAGCTGAGATTTAAAGCAGA 2124
DB 1256 RRR 1197
QY 2125 AAGCAAGAGATTACCAATTTGCACTGACAGATACAGATAGAGAGCTCAATTTCA 2184
DB 1196 RRR 1137
QY 2185 CAGCTTTCAGCTATTTGACCAAGCAAGAAAGACACCACTAGATTTGCAATTTGAGGCT 2244
DB 1136 RRR 1077
QY 2245 TAAAGACACCTTCTTCTGCGGACATCCAGGCC 2280
DB 1076 RRRRRRRRRATGCAAGCTCCCTCGACCTGACGCC 1041

RESULT 7
US-09-434-408-3
Sequence 3, Application US/09434408
Patent No. 6440697
GENERAL INFORMATION:
APPLICANT: Venezia, Domenick
APPLICANT: Grossmann, Angelika
TITLE OF INVENTION: RING FINGER PROTEIN ZAPOPS
FILE REFERENCE: 98-41
CURRENT APPLICATION NUMBER: US/09/434,408
CURRENT FILING DATE: 1999-11-04
EARLIER APPLICATION NUMBER: US 60/108,258
EARLIER FILING DATE: 1998-11-12
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 2169
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide sequence of zapop3
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(2169)
OTHER INFORMATION: n = A,T,C or G
US-09-434-408-3

Query Match	1.68;	Score 41;	DB 4;	Length 2169;
Best Local Similarity	21.88;	Pred. No. 0.045;		
Matches 111;	Conservative 115;	Mismatches 282;	Indels 1;	Gaps 1;

QY	1498	GAGCATCCGAGTAACACCACTTAATACCTACTGTAGTGTCTCCAGCTCCATCTTAA	1557
Db	670	GACRCARATYGGNATHGARBAAYMSNMONGAYMSNCCNGAIVGCMCAACNGAIVGNITTYSN	729
QY	1558	GCAAAATTACCTGTCCCTGGGAATAAAATTCTGTCACTTCAAGCATCTCCTACTGAGAAA	16177
Db	730	MGNGARCARYTTCNARTGGCRAAYMONTTYSNGAYTAAGAAABMNAARCARCARAAR	789
QY	1618	ATAGAATAAAAGAAAGATGTACATCATGCTCCGGATGAGATGAGATGACATTCATGATGTGA	167777
Db	790	ATGCTNGARCARYTTCNARTTTCGARMGNGNYTNGAIVTTGGCARMGNGARCAYACNCAR	849
QY	1678	CTTCTATGCGAGGGGTCAACCTTTAATGAAGAAAATGCCGTGATCTTAGCAACAACCTGTG	17377
Db	850	YTNTTNCARCARSMSMSMSCARAAARGAYGARATHYTNCARACNGTNAARCARBARCAR	909
QY	1738	AATTGGTGGCACATCATCTAGCTCATGTGAAGATGAACATTTCTTTTATGTGAGCTC	17977
Db	910	WSNMGNTTNCARARBSGNTTMSNGARCAYCARBMGCATTTTGAIVCGMARGMNCARBSG	969
QY	1798	TACAAAAGAGAATCTTAGACATTGGTAAAAAGCATGACATTACAGAACTTAATCTGTATG	1857
Db	970	YTNCARCARCARYTNAARCARACNGARCARAATAATHTMSMSNMGNATHCARABRYTNTN	10299
QY	1858	CTTGGAACCTGATCTCCCAAGCAACACAGGAAGACTACGAGGCTTCTAGAA-AAACTG	1916
Db	1030	CARATYAAYCARMGNCARARAARAAARMSMNSGAAATHYTTAAWSYTTGARARAAYARGN	10899
QY	1917	ACTGCAATTCCTCAGCATGCAATAGACTACTTTCACAGGCAAGAAATTAATCTCTGTCT	19767
Db	1090	ATHMGATGTGARCARYTNAATGMSNATHACNCAARGAAGAACGARMSTNTMNGMNGMNGN	11499
QY	1977	AGTSATTCACGAGTGCACAGCTCAAAATTTCT	2005
Db	1150	GAYGTGCMNSMNCMTGCAACRCARATGYT	1178

RESULT 8
 US-08-257-073-4
 : Sequence 4 Application US/08257073
 : Patent No. 5766597
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Paolelli, Enzo
 : APPLICANT: de Taisne, Charles
 : APPLICANT: Tine, John A.
 : TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
 : NUMBER OF SEQUENCES: 143
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Curtis, Morris & Safford, P.C.
 : STREET: 550 Fifth Avenue, 25th Floor
 : CITY: New York
 : STATE: New York
 : COUNTRY: UNITED STATES OF AMERICA
 : ZIP: 10036
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/257,073
 : FILING DATE: 09-JUN-1994
 : CLASSIFICATION: 424
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/075,783
 : FILING DATE: 11-JUN-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/852,305
 : FILING DATE: 18-MAR-1992

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/672,183
: FILING DATE: 20-MAR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Frommer, William S.
: REGISTRATION NUMBER: 25,506
: REFERENCE/DOCKET NUMBER: 454310-2570
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 840-3333
: TELEFAX: (212) 840-0712
: TELEX: 425066 CURTMS
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2223 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-257-073-4

```

Query Match	1.4%	Score 37	DB 1	Length 2223
Best Local Similarity	48.1%	Pred. No	0.75	
Matches 137	Conservative	0	Mismatches 145	Indels 3
				Gaps 1

Qy	1575	GGATATAAATTCGTGACTTCAAGATGTCCTACACAAAAATAGATATAAGAGAT	1630
Db	1528	GAATATATATACACTTAATGTTGACCAACACTCAATCAAAAAAAAAACAAAAAT	1583
Qy	1635	GTAAC---ATCATGCTTCCGAGATGAGAGATGACATCAATGATGTGACTTATGCGAGG	1691
Db	1588	GAACTATCTGATCTGATAGATGAGAAAAATTTTGATTAATCATCTGAAAAATATTATTTAAAGA	1644
Qy	1692	GTCACCTTATAGAAAGAAATGCGTCGATCTTAGCAACAACTCGAATGGTTGGACA	1750
Db	1648	GAATATTATTATGATGAGAAATGATATAGGAAGTAAAGTTTAAAAAAATATAGTGTGACA	1707
Qy	1752	CTCATTCAGTCATATAAGATGGAACCATTTCTTTATATGAGAGCTCTACAAAAGAAATC	1811
Db	1708	TTAAAAAAATTTGACACCTTAAAAAATGGAATAGTAGAACAATTAAATGATTCAT	1767
Qy	1812	TTACACATGTGTTAAAAAGCATGACATTTACAGAACTTAACCTGAT	1856
Db	1768	TTTAGAATAAAGTATAAAAAACCATGAAGCTATATAAACAGAT	1812

```

RESULT 9
: US-09-122-400B-5/c
: Sequence 5, Application US/09122400B
: Patent No. 6245974
: GENERAL INFORMATION:
: APPLICANT: Michalowski, Susan
: APPLICANT: Spiker, Steven
: TITLE OF INVENTION: MATRIX ATTACHMENT REGIONS
: FILE REFERENCE: Michalowski and Spiker
: CURRENT APPLICATION NUMBER: US/09/122,400B
: CURRENT FILING DATE: 1998-07-24
: PRIOR APPLICATION NUMBER: 60/066,118
: PRIOR FILING DATE: 1997-08-06
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 5
: LENGTH: 998
: TYPE: DNA
: ORGANISM: Nicotiana tabacum
: US-09-122-400B-5

```

	Query Match	1.4%	Score 36;	DB 4;	Length 998;
	Best Local Similarity	51.2%;	Pred.	No. 0.9;	
	Matches	84; Conservative	0; Mismatches	80; Indels	0; Gaps
OY	1997 CAAATTTCTTGAAAAAGCGTGAATCAATTGACAGACGAAAAGGATTTGGAGAANAAGACA	2056			
Db	809 CAAATTTACTGTAATAAAGCAAAAAAAAAACAAAATATATGTTCACAGTTTTCACAAAAAA	750			

[illegible]

```

RESULT 10
US-09-253-691-3
: Sequence 3, Application US/09253691
: Patent No. 6124100
: GENERAL INFORMATION:
: APPLICANT: Dong Kyu JIN
: TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
: FILE REFERENCE: 1942/36
: CURRENT APPLICATION NUMBER: US/09/253, 691
: CURRENT FILING DATE: 1999-02-22
: EARLIER APPLICATION NUMBER: KR 98-6, 278
: EARLIER FILING DATE: 1996-02-26
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: wordperfect 6.1/Windows
: SEQ ID NO 3
: LENGTH: 397
: TYPE: DNA
: ORGANISM: human
: US-09-253-691-3

```

[illegible]

RESULT 11
 US-08-700-651-1
 : Sequence 1, Application US/08700651B
 : Patent No. 6015882
 : GENERAL INFORMATION:
 : APPLICANT: PETERSEN, CAROLYN
 : APPLICANT: LEECH, JAMES
 : APPLICANT: NELSON, RICHARD, C.
 : APPLICANT: GUT, JIRI
 : TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
 : TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
 : TITLE OF INVENTION: INFECTIONS
 : FILE REFERENCE: 480.19-4(HV)
 : CURRENT APPLICATION NUMBER: US/08/700,651B
 : CURRENT FILING DATE: 1997-08-14
 : EARLIER APPLICATION NUMBER: 08/415,751
 : EARLIER FILING DATE: 1995-04-03
 : NUMBER OF SEQ ID NOS: 15
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 1
 : LENGTH: 5163
 : TYPE: DNA

US-08-700-651-1

[illegible]

```

1      RESULT 12
2      US-08-928-361B-4
3      ; Sequence 4, Application US/08928361B
4      ; Patent No. 6071518
5      ;
6      ; GENERAL INFORMATION:
7      ; APPLICANT: Peteresen, Carolyn
8      ; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
9      ; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
10     ; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
11     ; TITLE OF INVENTION: SPECIES INFECTIONS
12     ; NUMBER OF SEQUENCES: 30
13     ;
14     ; CORRESPONDENCE ADDRESSES:
15     ; ADDRESSEE: PETERS, VERNY, JONES & BIRKSA
16     ; STREET: 385 Sherman Avenue, Suite 6
17     ; CITY: Palo Alto
18     ; STATE: CA
19     ;
20     ; COUNTRY: USA
21     ; ZIP: 94306-1840
22     ;
23     ; COMPUTER READABLE FORM:
24     ; MEDIUM TYPE: Floppy disk
25     ; COMPUTER: IBM PC compatible
26     ; OPERATING SYSTEM: PC-DOS/MS-DOS
27     ; SOFTWARE: Patent Release #1.0, Version #1.30
28     ;
29     ; CURRENT APPLICATION DATA:
30     ; APPLICATION NUMBER: US/08/928,361B
31     ; FILING DATE: 12-SEP-1997
32     ;
33     ; CLASSIFICATION:
34     ; PRIOR APPLICATION DATA:
35     ; APPLICATION NUMBER: US 60/026,062
36     ; FILING DATE: 13-SEP-1996
37     ; ATTORNEY/AGENT INFORMATION:
38     ; NAME: Verny, Hana
39     ; REGISTRATION NUMBER: 30,518
40     ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
41     ; TELECOMMUNICATION INFORMATION:
42     ; TELEPHONE: 650-324-1677
43     ; TELEFAX: 650-324-1678
44     ;
45     ; INFORMATION FOR SEQ ID NO: 4:
46     ; SEQUENCE CHARACTERISTICS:
47     ; LENGTH: 5163 base pairs
48     ; TYPE: nucleic acid
49     ; STRANDEDNESS: double
50     ; TOPOLOGY: linear
51     ;
52     ; MOLECULE TYPE: DNA (genomic)
53     ;
54     ; US-08-928-361B-4

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:04:35 ; Search time 126 Seconds
(without alignments)
10340.079 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558
Sequence: 1 gggaccctgttgacccaagt.....cttataactcttacctat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135.8	5.3	2307	9	US-09-893-519A-87
2	38.6	1.5	5101	10	US-09-070-927A-516
3	37.8	1.5	438	10	US-09-960-352-11462
4	37.8	1.5	442	10	US-09-960-352-11227
5	36.8	1.4	579	10	US-09-864-761-9944
6	36.6	1.4	174424	10	US-09-967-768A-314
7	36	1.4	486	10	US-09-864-761-10113
8	36	1.4	510	10	US-09-864-761-18737
9	36	1.4	591	10	US-09-864-761-12567
10	36	1.4	998	10	US-09-816-894-5
11	36	1.4	2391	10	US-09-815-242-3827
12	36	1.4	2397	10	US-09-815-242-6667
13	35.8	1.4	1668	10	US-09-974-300-101
14	35.8	1.4	2534	10	US-09-764-878-403
15	35.6	1.4	640681	9	US-09-790-988-1
16	35.4	1.4	30350	9	US-10-118-328-3
17	34.8	1.4	1225	9	US-09-832-846-198
18	34.6	1.4	294	10	US-09-864-761-28655
19	34.6	1.4	561	10	US-09-864-761-12075

20	34.4	1.3	520	10	US-09-917-800A-1015	Sequence 1015, Ap
21	34.4	1.3	1137	9	US-09-815-242-9348	Sequence 9348, Ap
22	34.4	1.3	1278	10	US-09-764-868-280	Sequence 280, App
23	34.4	1.3	10249	12	US-10-002-600-65	Sequence 65, App
24	34.4	1.3	10249	12	US-10-044-090-5	Sequence 5, App1
25	34.4	1.3	57130	10	US-09-835-081-3	Sequence 3, App1
26	34.2	1.3	418	10	US-09-960-352-7989	Sequence 7989, Ap
27	34.2	1.3	1008	9	US-09-922-469-1	Sequence 1, App1
28	34.2	1.3	1008	10	US-09-923-236-1	Sequence 1, App1
29	34.2	1.3	1008	10	US-09-923-236-1	Sequence 1, App1
30	34.2	1.3	1872	10	US-09-815-242-4585	Sequence 4585, Ap
31	34.2	1.3	1878	10	US-09-815-242-8606	Sequence 8606, Ap
32	34.2	1.3	2338	9	US-09-286-488-13	Sequence 13, App1
33	34.2	1.3	2338	10	US-09-737-178-13	Sequence 13, App1
34	34.2	1.3	11854	10	US-09-070-927A-306	Sequence 306, App
35	34	1.3	385	10	US-09-960-352-1739	Sequence 1739, Ap
36	34	1.3	519	10	US-09-878-574-4292	Sequence 4292, Ap
37	34	1.3	3666	10	US-09-137-531-13	Sequence 13, App1
38	34	1.3	3666	10	US-09-137-531-14	Sequence 14, App1
39	34	1.3	4197	10	US-09-137-531-7	Sequence 7, App1
40	34	1.3	4197	10	US-09-137-531-8	Sequence 8, App1
41	34	1.3	13865	10	US-09-070-927A-200	Sequence 200, App
42	33.8	1.3	1930	12	US-10-003-152-15	Sequence 15, App1
43	33.8	1.3	2056	12	US-10-003-152-23	Sequence 23, App1
44	33.8	1.3	2056	10	US-09-809-790-3	Sequence 3, App1
45	33.8	1.3	2088	10	US-09-809-617-3	Sequence 3, App1

ALIGNMENTS

RESULT 1
US-09-893-519A-87
Sequence 87, Application US/09893519A
Publication No. US20030027243A1
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESTIVA, Thameara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARITSKY, Svetlana
APPLICANT: MENDILLO, Marc
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/IG548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-06-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 87
LENGTH: 2307
TYPE: DNA
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/Y11354.1
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(2307)
US-09-893-519A-87
Query Match 5.3%; Score 135.8; DB 9; Length 2307;

Db 51822 ATCTTACGAGTATGAGAAATGCAAAATATATATGGAATTTGGAAATTCACCAAAACACAC 51881
QY 308 TAATCAAGAAAGTGGCAGTGCACCTGTGTAAANA 342
Db 51882 TGATTAGAGAAAAGAACTACTATATATGAGAAA 51916

RESULT 7
US-09-864-761-10113
; Sequence 10113, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10113
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL035457.11
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
US-09-864-761-10113

Query Match 1.4%; Score 36; DB 10; Length 486;

Best Local Similarity 45.4%; Pred. No. 2.9; Indels 0; Gaps 0;
Matches 129; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 968 AGAAGTGTGTGTCAGTCAAACTTTGMAACCACTTGCTGCTCAGTGGAGCAAAAGCTGGA 1027
Db 179 AGCTCAGAGGCGAGGAGATATCAACACCTGCGCTCAGACATCTGTATAGCAGCAATACCT 238
QY 1028 GTTGTGACACTTCAATTTGTGTGGGCCCAACTGTGCAACAGAGAGGAGCAACAGCTGGAAT 1087
Db 239 GCTTCAGCACCCTGCATCAGATCAACACCCGCTCAACATCTGCATCAACATCTGCACCA 298
QY 1088 GGTGTTGCTGAGCTTCAAAACCACTGTGACATCTGTGCAAAACAGTGCACGCTC 1147
Db 299 GCATCACTACAGCATGAGATCACTACACAGCATCAGCATCTGCATCTGCATCAACGAT 358
QY 1148 TCACTGCACTGAAAGCCAGTGTCTGTGAAACAGCACTGATCCCTTCAGCA 1207
Db 359 TTTTCAGCACCCTGATCAACACCTGCTTCAACATCTGCATCAACATCTGCACAAAGCTCA 418
QY 1208 GTAACTTTGGAGAACTTCAAGTGCAGCTATTTGTCTTCATC 1251
Db 419 CTACCATCAGCATCCTACACAGCATCAGCATCTGCACCTGCATC 462

RESULT 8
US-09-864-761-18737
; Sequence 18737, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 18737
LENGTH: 510
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL035457.11
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.85
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.95
OTHER INFORMATION: NT HIT: U01287.1, EVALUATE 1.30e+00
US-09-864-761-18737

Query Match 1.4%; Score 36; DB 10; Length 510;
Best Local Similarity 45.4%; Pred. No. 3;
Matches 129; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 968 AGACTGTGTGAGTGAACCTTTGAACCCACTTGTCTGTCACAGTGGACCAAAAGCTGA 1027
DB 42 AGCTAGAGGCCAGCAGCATCATCAACCTGCTCAGCATGTGTATCAGCACCAATACCT 101
QY 1028 GTTGTGACACTTCTGTGGGCCCACTGCTGCAACAGGAAACACAGCTGGAACT 1087
DB 102 GCTTACAGCACCCTGATCAGCATCAACACCCGCTTCAACATCTGCATCAACATCTGCACCA 161
QY 1088 GGTTCCTTACAGCTTCAAAACCACTTGTGACATCTGTGCAACACAGTGAACAGGTC 1147
DB 162 GCATCATCTACAGCATCATCAGCATCAGCATGAGCATCTGCACTTGCATCATCAACAT 221
QY 1148 TCACGTGACCTGGAAGCACTTGTCTGTGGAACAGCATGTAACACTGTCCTTCCAGCA 1207
DB 222 TTTTCAGCACCCTGATCAACACCTGCTTCAACATCTGCTATCAACATCTGCAACAGCATCA 281
QY 1208 GTAACTTTTGGAAACTTCAAGTGCAGCTATTGTCTTCATC 1251
DB 282 CTACCATCAGCATCATCAGCATCAGCATGAGCATCTGCACCTGCATC 325

RESULT 9
US-09-864-761-12567
Sequence 12567, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12567
LENGTH: 591
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC019141.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.85
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.62
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.65
US-09-864-761-12567

Query Match 1.4%; Score 36; DB 10; Length 591;
Best Local Similarity 50.6%; Pred. No. 3.3;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 1614 AAAATAGATTAAGACATGTAACATCATCTCCGAGATGAGATGATCATATGAT 1673
DB 233 AGAAATAGAGTAGAGAAACAAACAGCACCAAGAAACATGACATGAAATGGGAT 292
QY 1674 GTGACTTCTATGAGGAGGCTACCTTAATGAAGAAATGCGCTGATCTTACCAACAAC 1733
DB 293 GTGCTTACAGTACGTGGGTTAACCATTAACATATCTTTGTGGGATGATTAATAATAC 352
QY 1734 TCTGATTGGTTGGCACAACCTCATTCAGTCATGTAAGATGAACCATTTCTTT 1785
DB 353 TCTCAAGTAGATGATVACCTGTTACCTTCTTACGTGAATACAGTTTATTT 404

RESULT 10
US-09-816-894-5/c
Sequence 5, Application US/09816894
Patent No. US20020073448A1
GENERAL INFORMATION:
APPLICANT: Michalowski, Susan
APPLICANT: Spiker, Steven
TITLE OF INVENTION: MATRIX ATTACHMENT REGIONS
FILE REFERENCE: 5051.401XXDV
CURRENT APPLICATION NUMBER: US/09/816,894
PRIOR FILING DATE: 2001-06-20
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 998
TYPE: DNA
ORGANISM: Nicotiana tabacum
US-09-816-894-5

Query Match	36	1.4%	Score	36	DB	10	Length	998	
Best Local Similarity	51.2%		Pred. No.	4.6					
Matches	84	Conservative	0	Mismatches	80	Indels	0	Gaps	0
QY	1997	CAAAATTTCTGAAAAAGCTGGATCAATTGTGAGAGACAGACAGAAAGATTGTGAGAAAGACAGA	2056						
Db	809	CAAAATTTCTGAAAAAGCTGGATCAATTGTGAGAGACAGACAGAAAGATTGTGAGAAAGACAGA	750						
QY	2057	AATGTTACTTAAAGCAGCCAGAGTCGTTCTAATTAAGAAAGATCCAGAACACTGAGATT	2116						
Db	749	AACCTCTTTAAAGAACTGAAAAAATTTTTCCTAAAAATAATTTTGTAAAAAACTGAAAAA	690						
QY	2117	AAAGCAGAAAGCCAGAGATTACAGACATTGGACCTTGACAGA	2160						
Db	689	AAAGCAGAAAGCTAATTTTCTTAAGTAATTAAAAAACTGAAAAA	646						

```

US-09-815-242-3827
RESULT 11
Sequence 3827, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELTRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3827
LENGTH: 2391
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-815-242-3827

```

	Query Match	Best Local Similarity	1.4% 51.9%	Score 36: Pred. No. 8.2:	DB 10:	Length 2291:
	Matches	81:	Conservative	0:	Mismatches	75: Indels
					0:	Gaps
QY	189	ATCTCTCAGCAAACTGTATACAGAGCCGAGACCAAGTAACATATACCTCAAGGCCAGC	248			
Db	346	AGCAACACACAAAACCGTAGCAATCAAAACAATATACAAACAATCAAAACCGCAATACCCAA	405			
QY	249	AGTACACAGCAATCTCTCAAAACAGTCAAAATGTGTACAGTGCAGAACTAGCTCACAATT	308			
Db	406	AACCATTAACGGTTCTTACACAGCAATCAAAATCGACACGCCAAAAATACAAATGCGCGGAAC	465			
QY	309	AATCAAGAAAGTGGCAGTGCACCTCGTTAAAAAATT	344			
Db	466	AACCAAAACCGTGGCGGCCAAACCGTAAACCAAT	501			

```

RESULT 12
US-09-815-242-6667
Sequence 6667, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 6667
LENGTH: 2397
TYPE: DNA
ORGANISM: Enterococcus faecalis
FEATURES:
NAME/KEY: CDS
LOCATION: (1)..(2397)
US-09-815-242-6667

```

```

Query Match      1.4%: Score 36; DB 10; Length 2397;
Best Local Similarity 51.9%; Pred. No. 8.2;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 189 ATCTCTCAGCAGCAACTGTAAACAAGCCGAGACACAGTAACATATACCTCAAGGCCAGC 248
Db 346 AGCAACACACACAACAAACCGTAGCAATCAACCAAAATACAAACATCAAAACCGCAATACCCA 405

QY 249 AGTACACACGAATCCTCAAAACAGTCAAAAATCTGTACAGTGGCCGAACCTTAGCTCACAATT 308
Db 406 AACATTAACGGTTCTACACAACGAATCAAAATCGGACACAGCCAAAATAATMAACAATGGCGGAAC 465

QY 309 AATCAAGAAAGTGCGAGTGACACTGTTAAAAAATT 344
Db 466 AACCAAAACCGTGGCGGCCAAACCGTAAACAACAATT 501

RESULT 13
US-09-974-300-101
: Sequence 101, Application US/09974300
: Patent No. US20020146721A1
: GENERAL INFORMATION:
: APPLICANT: Berka, Randy M.
: APPLICANT: Clausen, Ib Groth
: TITLE OF INVENTION: Methods For Monitoring Multiple Gene
: FILE REFERENCE: 10085,500-US
: CURRENT APPLICATION NUMBER: US/09/974,300
: CURRENT FILING DATE: 2001-10-05

```

Db 265828 AA 265829

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:39:25 ; Search time 3476 Seconds

(without alignments)
11918.313 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558

Sequence: 1 gggacccttggtaccacaaagt.....cttattacttaccat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues 32308132

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estda:*
2: em_esthum:*
3: em_estlin:*
4: em_estlu:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlum:*
16: em_estlom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_hiv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	634.2	24.8	946	14	BQ952832
2	537.6	21.0	673	13	BI517527
3	398	15.6	1116	11	AK012135
4	395.2	15.4	599	10	BB616334
5	370.6	14.5	388	9	AA457205
6	344	13.4	427	10	AW297262

7	286.4	11.6	570	13	BM539135	BM539135 hb05c07.g
8	288.6	11.3	1410	14	BQ220171	BQ220171 AGENCOURT
9	286.4	11.2	787	12	BG213757	BG213757 RST33376
10	269.4	10.5	2046	11	BC026493	BC026493 Mus muscu
11	267.8	10.5	777	12	BG217524	BG217524 RST37234
12	262.4	10.3	507	14	BQ560074	BQ560074 H4062B07-
13	248.8	9.7	545	10	AW390057	AW390057 RGS-ST017
14	244	9.5	600	13	BI986567	BI986567 3173-36 M
15	241	9.4	734	12	BG259865	BG259865 602933377
16	238.8	9.3	940	13	BI152593	BI152593 602918459
17	227.6	8.9	404	14	BQ417829	BQ417829 1K51007.Y
18	226.2	8.8	417	9	AL121204	AL121204 DKFZP762C
19	222.8	8.7	884	9	AL668344	AL668344 AL668344
20	208.8	8.2	552	10	AW964632	AW964632 EST376705
21	208.6	8.0	271	12	BG191106	BG191106 RST10076
22	203.2	7.9	233	12	BG198225	BG198225 RST17369
23	184	7.2	556	14	BQ560102	BQ560102 H4062C09-
24	183.6	7.2	518	9	AA438072	AA438072 vd22a01.s
25	183	7.2	454	12	BE939960	BE939960 RC1-UT003
26	178.8	7.0	563	12	BE966487	BE966487 UT-M-CGP
27	178.2	7.0	629	14	BM938537	BM938537 UT-M-CGP
28	171.4	6.7	469	12	BE698723	BE698723 RC1-UT003
29	165.4	6.5	463	9	AA549039	AA549039 VK77901.s
30	164	6.4	410	9	AI503691	AI503691 VK77901.x
31	161.6	6.3	541	12	BG514288	BG514288 dc96d09.x
32	161.4	6.3	749	14	BM853763	BM853763 K-EST0135
33	158.8	6.2	409	12	BE135081	BE135081 601780119
34	158.2	6.2	468	10	AW785378	AW785378 116368 MA
35	154.2	6.0	484	12	BE940003	BE940003 RC1-UT003
36	151.2	5.9	623	10	AV965961	AV965961 AV965961
37	148.4	5.8	686	17	AG092765	AG092765 Pan trogl
38	143.2	5.6	355	10	AW637728	AW637728 b161h05.v
39	142	5.6	426	10	BE146234	BE146234 MRO-HT020
40	141.2	5.5	331	9	AI467327	AI467327 vd22a01.x
41	135	5.3	919	14	BQ685051	BQ685051 AGENCOURT
42	134.2	5.2	384	10	AW378648	AW378648 PMO-HT022
43	134	5.2	530	12	BE844520	BE844520 EST265 AP
44	130.2	5.1	312	10	AW414768	AW414768 48566 MAR
45	128	5.0	282	9	AA360405	AA360405 EST69559

ALIGNMENTS

RESULT 1
BQ952832
LOCUS BQ952832
DEFINITION AGENCOURT_8949744 NCI_CGAP_Mam2 Mus musculus CDNA clone
IMAGE:6442204 5', mRNA sequence.
ACCESSION BQ952832
VERSION BQ952832.1 GI:22368310
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 946)
NITH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
<http://image.jnl.gov>
Plate: LAM3966 row: k column: 05
High quality sequence stop: 666.

FEATURES
SOURCE location/Qualifiers
1..946

```

/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:6442204"
/clone_1lb="NCI CGAP Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: Not; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT      322 a      203 c      193 g      226 t      2 others
ORIGIN

Query Match      24.8%; Score 634.2; DB 14; Length 946;
Best Local Similarity 85.0%; Pred. No. 1.9e-169;
Matches 764; Conservative 0; Mismatches 110; Indels 25; Gaps 4;

OY 1409 GTAGTTACAGCCTTCAGAGGCAATGAAAAAGTACCAATTCACATTCCTCA 1468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35 GTGTACACAGCCTTCAGAGGCAATGAAAAAGTACCAATTCACATTCCTCA 94

OY 1469 ACATTGACATTCAGAAATGTGTGAGACAGAGATGCGATGAAACCATATACCTACT 1528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 CCATTGACACCTCAGAACTGTGGGACAGAAAGAC---ACCAATTAATGCTGTGATGCTTACT 151

OY 1529 ACATGACTTTCCTCAGCTTCATCTTAAAGCAATTAACCTGCTGAAATTAATTC 1588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 A-----CTTCATCATTAAGCAAAATTA-CTGTGCTGGAATTAACCTTT 195

OY 1589 GTTCATTCAGC-----ATCTCTACTCAGAAAAATAGAAATTAAGAAATTAACATC 1642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 GTGCTTCAGCTCAAGTAAAGATCTTCTTATTCAGACCAATTAATTAAGAAATGAGACCAAC 255

OY 1643 ATGCTCCGAGATGAGATGACATCAATGATGATCTTATGAGGAGGCTCAACCTTAA 1702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 ATGCTTCAGAGTGAAGATGACATCAATGATGATCTTATGAGGAGGCTCAACCTTGA 315

OY 1703 TGAGAAAAATGCTGATCTTAAACAACAACCTGTAATGTTGGACACTCATTCAGTC 1762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 TGAAGAAATGCTGATCTTAAACAACAACCTGTAATGTTGGACACTCATTCAGTC 375

OY 1763 ATGTAAGATGAACATCTTCTTATTAAGAGCTTACAAAAGAGAACTTAAACATTCG 1822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 ATGTAAGATGAACATCTTCTTATTAAGAGCTTACAAAAGAGAACTTAAACATTCG 435

OY 1823 TAAAAAGCATGACATTAAGAACTTAACTGCTGTAATGCTGTAATGCTCCCAAGCAAC 1882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 TAAAAAGCATGACATTAAGAACTTAACTGCTGTAATGCTGTAATGCTCCCAAGCAAC 495

OY 1883 ACAGGAGACATTAAGAGGCTTCTAGAAAACTGACTGCAATTCCTCAGCATGAAATGAC 1942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 ACAGGAGACATTAAGAGGCTTCTAGAAAACTGACTGCAATTCCTCAGCATGAAATGAC 555

OY 1943 TACTTCAAGGGAAGTAAATTAACATCCGTTAGTATACAGCTCCACAGCTCAAAAT 2002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 TACTTCAAGGGAAGTAAATTAACATCCGTTAGTATACAGCTCCACAGCTCAAAAT 615

OY 2003 TCTTGAAGAGCTGATCAATTTGAGAGCAGAGAAAGATTTGGAAGAAAGAGAAATGTT 2062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 TCTTGAAGAGCTGATCAATTTGAGAGCAGAGAAAGATTTGGAAGAAAGAGAAATGTT 675

OY 2063 ACTTAAAGCAGCCAGAGTCTTCTAATAAAGAGATCCAGAAACAGCTGAGATTAAGCA 2122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 ACTTAAAGCAGCCAGAGTCTTCTAATAAAGAGATCCAGAAACAGCTGAGATTAAGCA 735

OY 2123 GAAAGCAAGAGTTACACAAATTTGAGAGATTCAGACATTAAGAGAGAGCTTAATCT 2182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 736 AAAAGCAAGAGTTACACAAATTTGAGAGATTCAGACATTAAGAGAGAGCTTAATCT 795

OY 2183 CACAGCTCTTGAGATTTGAGACCAAGAAAGAGACCATTAAGATTTGAGAGG 2242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 796 CACAGCTCTTGACACTTATGAGCAAGAAAGAGACCACTAGATCAAGAAATGAGAG 855
OY 2243 CTTAAAGACACCTCTTCTTGAGACATTCAGCTGACAGCCACCAACAGTTGC 2301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 856 CTTCAAGATTAACCTTCTACTCTGAGACTTCAGAGCTTACAGCTTACAGCAACCAACATTTTC 914

RESULT 2
BI517527      673 bp      mRNA      linear      EST 29-AUG-2001
LOCUS        603041813F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5162978 5',
DEFINITION   mRNA sequence.
ACCESSION    BI517527
VERSION      BI517527.1 GI:15342319
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 673)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-remail.nih.gov
              Tissue Procurement: Life Technologies, Inc.
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: The I.M.A.G.E. Consortium (LLNU)
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNU, at:
              http://image.llnl.gov
              Plate: LLAM1404 row: j column: 03
              High quality sequence stop: 673.
              Location/Qualifiers
                1..673
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:5162978"
                  /clone_1lb="NIH_MGC_116"
                  /lab_host="DH10B"
                  /note="Organ: pooled colon, kidney, stomach; Vector:
                  pCMV-SPORT6; Site_1: Not; Site_2: EcorV (destroyed); RNA
                  source anonymous pool of 3 colons, age 26 yo male, 49 yo
                  female, 71 yo male colon; 46 yo male kidney, and pool of 2
                  stomachs, 62 yo male and 70 yo female. Library is
                  oligo-dt primed and directionally cloned (EcorV site is
                  destroyed upon cloning). Average insert size 1.4 kb,
                  insert size range 1-3 kb. Library is normalized and
                  enriched for full-length clones and was constructed by C.
                  Gruber (Invitrogen). Research Genetics tracking code
                  023. Note: this is a NIH_MGC Library."

BASE COUNT      224 a      146 c      137 g      166 t
ORIGIN

Query Match      21.0%; Score 537.6; DB 13; Length 673;
Best Local Similarity 98.1%; Pred. No. 6.3e-142;
Matches 565; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

OY 1985 CAGGTACAGCTCAAAATTTCTGAAGAGTCAATTCGAGAAACAGAGAAAGATTT 2044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CAGGTACAGCTCAAAATTTCTGAAGAGTCAATTCGAGAAACAGAGAAAGATTTG 60

OY 2045 GGAAGAAAGAAATGTTACTTAAGCAGCCAGAGTCTTCTAATAAAGAGATCCAGA 2104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GAAGCAAGAAAGAAATGTTACTTAAGCAGCCAGAGTCTTCTAATAAAGAGATCCAGA 120

OY 2105 ACAGCTGAGATTAAAGCAAGAAAGAGATTAAGCAATTTGAACTTGACAGATACA 2164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ACAGCTGAGATTAAAGCAAGAAAGAGATTAAGCAATTTGAACTTGACAGATACA 180

OY 2165 GCATTAGAGCGTAAATCTACAGCTCTTGAGATTTGAGACCAAGAAAGAGACACT 2224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 181 GCATGAGACGCTATCTCACAGCTCTTGACGCTATTGGACCAAGGAAGACACCACT 240

Oy 2225 AGAATCTGGAATTGAGGGCTTAAAAAGACACCTTCTGCTTCTGGAGATCCAGCCGTGAC 2284

Db 241 AGAATCTGGAATTGAGGGCTTAAAAAGACACCTTCTGCTTCTGGAGATCCAGCCGTGAC 300

Oy 2285 AGCCACCAAAAGTTGATGCTGCTCCAGAAATCACGGAATCTGCTCCAGGACTTATATT 2344

Db 301 AGCCACCAAAAGTTGATGCTGCTCCAGAAATCACGGAATCTGCTCCAGGACTTATATT 360

Oy 2345 TTGTATGAGACAGAGAGGAGATGATTTCTGAGCTCTATACCTGGGCCCTCTGAA 2404

Db 361 TTGTATGAGACAGAGAGGAGATGATTTCTGAGCTCTATACCTGGGCCCTCTGAA 420

Oy 2405 GTGACCACTCCACTCTTCCATCCACATCCTTGTCTGCTTACTGCCAAGACACAAAGC 2464

Db 421 GTGACCACTCCACTCTTCCATCCACATCCTTGTCTGCTTACTGCCAAGACACAAAGC 480

Oy 2465 ATGTGTCACCTGCTCTGAAATTTCAATTTCTGGAATAATTA-CACCAACATGGAAGAGCAT 2523

Db 481 ATGTGTCACCTGCTCTGAAATTTCAATTTCTGGAATAATTAACCAACATGGAAGAGCAT 540

Oy 2524 TGTTCACGATAG-AACCTTATTAACCTTACCTAT 2558

Db 541 TGTTCACGATAGAACTTATTAACCTTACCTAT 576

RESULT 3
AK012135

LOCUS AK012135 1116 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610524B04;homolog to TRANSCRIPTION INITIATION FACTOR TFIIID 105 KDA SUBUNIT (TAFII-105), (TAFII105), full insert sequence.

ACCESSION AK012135

VERSION AK012135.1 GI:12848693

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain: C57BL/6J) 10 days embryo cDNA to mRNA, clone_11b:RIKEN full-length enriched mouse cDNA library clone:2610524B04.

ORGANISM Mus musculus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE 3 Genome Res. 10 (11), 1757-1771 (2000)

TITLE Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, T., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Kashiwagi, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL MEDLINE 20530913

PUBMED 11076861

AUTHORS Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Atakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,

Atakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakawa, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pease, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mommaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, T., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seyer, T., Shibata, Y., Storch, K. F., Suzuki, H., Togo, O., Wang, K. H., Weitz, C., Whiteaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohlschl, S., and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)

JOURNAL MEDLINE 21085660

PUBMED 11217851

AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Atakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

Direct Submission Submitted (10-JUL-2000) Yoshinori Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

FEATURES

SOURCE

Location/Qualifiers

1..1116

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="PANTOM:DB:2610524B04"

/db_xref="MGD:MGI:1905330"

/db_xref="taxon:10090"

/clone="2610524B04"

/clone_11b="RIKEN full-length enriched mouse cDNA library"

/dev_stage="10 days embryo"

misc_feature

1..1116

/note="data source:SPRR, source key:Q92750, evidence:ISS homolog to TRANSCRIPTION INITIATION FACTOR TFIIID 105 KDA SUBUNIT (TAFII-105) (TAFII105)"

/db_xref="MGD:MGI:1919754"

BASE COUNT 216 a 347 c 330 g 223 t

ORIGIN

Query Match 15.6%; Score 398; DB 11; Length 1116;

Best Local Similarity 83.5%; Pred. No. 7.7e-102;

Matches 476; Conservative 0; Mismatches 90; Indels 4; Gaps 2;

Oy 1 GGGACCTGGTGGTCAAGGAGGCTGGGACGAGCCCTCAAGAGCAGCGGCTT 60

Db 551 GGGACCTGGTGGTCAAGGAGGCTGGGACGAGCCCTCAAGAGCAGCGGCTT 607

Oy 61 AGGCTGCTGCTCTCAGATAGTGCCTGGAAGGCCCAACACGACGACATCCAGTTT 120

Db 301 CTTCAGAGATACCCCTCTCTACTCTGGAGCTTCAGCCCTACAGCCACCAACCAATT--- 357
 Qy 2303 TCGTCCAGATATCAGAGAAATCGCTTCAGAGCTTGATATTTTATGAGAAGAGACG 2362
 Db 358 TCGTCCAGATATCAGAGAAATCGCTTCAGAGCTTGATATTTTATGAGAAGAGACG 417
 Qy 2363 GGAGATGAGATATTCGAGCTGTATACCTGGCCCTTCGTGAAGACACATCCACTTTC 2422
 Db 418 AGAGATGAGATATTCGAGCCCTATACCTTCGCCCTCTGTAAGATGATGATCA---TC 473
 Qy 2423 CATCCACATCTCTGCTATTTATCTGCCAAGAAGACAC--AAGCATTTTGTGACATGCTCT 2480
 Db 474 CATCCAGACGACACTGTTTACTGTCAGAAAGACAGACAGACACTGTTGACATGTTCT 533
 Qy 2481 GAAATTCATTT 2492
 Db 534 GAACCTTCCTT 545
 RESULT 5
 AA457205 388 bp mRNA linear EST 06-JUN-1997
 LOCUS aa88609.r1 Stragatene fetal retina 937202 Homo sapiens cDNA clone
 DEFINITION IMAGE:838408.5' similar to SW:TR11.DROME P47825 110 KD
 TBP-ASSOCIATED FACTOR ; mRNA sequence.
 ACCESSION AA457205 GI:2179925
 VERSION AA457205
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 388)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenger, K., Steptoe, M., Tan, F., Theisling, B., White, T., Wyllie,
 'J., Waterston, R. and Wilson, R.
 Washu-Merck EST Project 1997
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -28ml3 rev1 ET from Amersham.
 FEATURES
 source
 1. 388
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:838408"
 /clone_lib="Stratagene fetal retina 937202"
 /sex="mixed"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled
 retinal tissue. Average insert size: 1.0 kb. Uni-ZAP XR
 Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
 adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."
 BASE COUNT 117 a 88 c 77 g 106 t
 ORIGIN
 Query Match 14.5% Score 370.6; DB 9; Length 388;
 Best Local Similarity 98.7% Pred. No. 2.4e-94;
 Matches 384; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
 Qy 589 CAAAGACTCTCTGCAATGTAATAAACTAGCATGTAGTGCACAGTCCCTGGAAT 648
 Db 1 CAAAGACTCTCTGCAATGTAATAAACTAGCATGTAGTGCACAGTCCCTGGAAT 60
 Qy 649 GGGGCAAAATGTGAAGAGCTGTGGAACAACCTTTGGATGCAAAAATGGAAGAGAGA 708

Db 61 GGGTCTAATGTGAAGAGCTGTGGAACAACCTTTGGATGCAAAAATGGAAGAGAGA 120
 Qy 709 ATTACTGGAACCTGTATGTTGAACATCAAGCTTCACCTAGAGCTCAGCTGCTCTT 768
 Db 121 ATTACTGGAACCTGTATGTTGAACATCAAGCTTCACCTAGAGCTGCTCTT 180
 Qy 769 TCTTAAGAAAGCGTGTGCTTACGACAACTTGCTTAACCTCCAGAGCTTCATCA 828
 Db 181 TCTTAAGAAAGCTGTGCTTACGACAACTTGCTTAACCTCCAGAGCTTCATCA 240
 Qy 829 GCAATGTGTGACGACACTTGTAGTACATGGCTCATCTGCTACTACAAAGTAAAC 888
 Db 241 GCAATGTGTGACGACACTTGTAGTACATGGCTCATCTGCTACTACAAAGTAAAC 300
 Qy 889 AACCTTCCTGCTGTGACAACTACAGTGTCCCAAGCCAGCTGTGAAGAGTCAATATGT 948
 Db 301 AACCTTCCTGCTGTGACAACTACAGTGTCCCAAGCCAGCTGTGAAGAGTCAATATGT 360
 Qy 949 TTCTGAGCAACAGCAGCCAGAACTGTGT 977
 Db 361 TTCTGAGCAACAG-ACCAGAACTGTGT 388
 RESULT 6
 AM297262/C 427 bp mRNA linear EST 16-JAN-2000
 LOCUS UT-H-BM0-aj3-c-11-0-UI.s1 NCI_CGAP_Sub6 Homo sapiens cDNA clone
 DEFINITION IMAGE:2731796.3' mRNA sequence.
 ACCESSION AM297262
 VERSION AM297262
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 427)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first A
 tail. cDNA library preparation: M.B. Soares Lab Clone distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LNL at:
 www-bio.lnl.gov/dbp/image/image.html
 Seq primer: M13 Forward
 POLYA=yes
 FEATURES
 source
 1. 427
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2731796"
 /clone_lib="NCI CGAP Sub6"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT30-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub6
 is a subcloned library derived from BW, which consists of
 a mixture of four normalized libraries: NCI_CGAP Brn50,
 NCI_CGAP Lw13, NCI_CGAP_Ov18, G8C1. The NCI_CGAP_Sub6
 library had 7 million recombinants. A single-stranded DNA
 preparation of BW was used as a tracer in a subclative
 hybridization with a driver comprising: the IMAGE pool
 (NCI_CGAP_Kid3 pool 1 LHAM 3334-3337, 3682-3683,
 3798-3803 (IMAGE Clonoids 1322376-1323911,
 1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1
 LHAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids
 1323912-1325831, 1471368-1472903, 1492104-1493255);
 NCI_CGAP_Lw5 pool 1 LHAM 3575-3582, 3851-3854 (IMAGE

Clonoids 1414920-1417991, 1520904-1524391: NCI CGAP_GC4
 pool 1 LHAM 3164-3167 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631, 1469064-1470963, 1475592-1476743); NCI CGAP_P122 pool 1 LHAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959, 1217928-1220615); NCI CGAP_C010 pool 1 LHAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351) (30% of the driver population), plus a pool of 3,840 arrayed clones from NCI CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and NCI CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (20% of the driver population), plus a pool of 11,136 clones from NCI CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (30% of the driver population).
 Subtraction was performed as previously described [Ronald L. Lennan & Soares (1996): Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6, 791-806].
 TAG_L1B=Cbcl
 TAG_TISSUE=B
 TAG_SEQ=TCA
 BASE COUNT 109 a 75 c 111 g 132 t
 ORIGIN

Query Match 13.4%; Score 344; DB 10; Length 427;
 Best Local Similarity 92.6%; Pred. No. 1e-86;

Matches 411; Conservative 0; Mismatches 15; Indels 18; Gaps 4;

QY 1186 AGTACACTGTCCTCCAGCAGTAACTTTGGAGAACTTCAGTGACAGTATTGCTC 1245
 DB 427 AGTACACTGTCCTCCAGCAGTAACTTTGGAGAACTTCAGTGACAGTATTGCTC 368
 QY 1246 TCATCTGTGAACCTGTTGCTTCCTTCCTGCTGGACACATCTG-CAAGCTGTTATG 1304
 DB 367 TCATCTGTGAACCTGTTGCTTCCTTCCTGCTGGACACATCTG-CAAGCTGTTATG 309
 QY 1305 GGACTCCAGTCAATCAACTTCCAGCGCGGCGCTGCTTTACACACAGCTGGGA 1364
 DB 308 GGACTCCAGTCAATCAACTTCCAGCGCGGCGCTGCTTTACACACAGCTGGGA 249
 QY 1365 TTCCAAACAGGCACTTCAAGACCAACTATTCTGATGTTTCAAGTTCAGCAGCTT 1424
 DB 248 TTCCAAACAGGCACTTCAAGACCAACTATTCTGATGTTTCAAGTTCAGCAGCTT 204
 QY 1425 CAGAGGCAATGAAAAAAGTAGACACAAATTCACATTCCTCAACATTGACATTGACA 1484
 DB 203 CAGAGGCAATGAAAAAAGTAGACACAAATTCACATTCCTCAACATTGACATTGACA 144
 QY 1485 AATGTGACAGAGAGATGCCAGTGAACACATTAATACCTACTAGTCAAGTTCCTCAG 1544
 DB 143 AATGTGACAGAGAGATGCCAGTGAACACATTAATACCTACTAGTCAAGTTCCTCAG 84
 QY 1545 CTTCATTTCTAAAGCAATTAATCTGCTGGAATTAATTTGTGACTTCAAGCATCT 1604
 DB 83 CTTCATTTCTAAAGCAATTAATCTGCTGGAATTAATTTGTGACTTCAAGCATCT 25
 QY 1605 CCTACTCAGAAAAATGAAATTAATA 1628
 DB 24 CCTACTCAGAAAAATGAAATTAATA 1

RESULT 7
 BM539135 570 bp mRNA linear EST 20-FEB-2002
 LOCUS hb05c07.g1 Canis cdnas from testes cells Canis familiaris cdna
 DEFINITION Clone hb05c07 5', mRNA sequence.
 ACCESSION BM539135
 VERSION BM539135.1 GI:18820807
 KEYWORDS EST.
 SOURCE dog.
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 REFERENCE 1 (bases 1 to 570)

AUTHORS O'Shaughnessy, A.L., McComble, W.R., Baker, J.P., Balija, V., Cummins, D., Dedhia, N.N., de la Bastide, M., Katzenberger, F., King, L., Kitchoff, K.A., Miller, B., Muller, S., Nascento, L.V., Palmer, L., Santos, L., Shah, R.S., Spiegel, L.A., Zuluvern, T., Preston, R. and Hanon, G.J.
 TITLE
 JOURNAL
 COMMENT
 Unpublished (2002)
 Contact: W. Richard McComble
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884
 Fax: 516 367 8874
 Email: mcomble@cshl.org
 Plate: hb05 row: c column: 07
 Seq primer: -21M3univRev
 High quality sequence stop: 570.
 location/Qualifiers
 1. 570
 /organism="Canis familiaris"
 /db_xref="taxon:9615"
 /clone="hb05c07"
 /tissue_type="testes"
 /note="Vector: Lambda Zap II; The library was produced by Greg Hanon and Raymond Preston (Cold Spring Harbor Laboratory). This library is oligo(dT) primed using stratagene zap cdna synthesis kit. It was made from dog testes. Please contact Greg Hanon (hanon@cshl.org) with any library related inquiries."

BASE COUNT 168 a 116 c 123 g 163 t
 ORIGIN

Query Match 11.6%; Score 296.4; DB 13; Length 570;
 Best Local Similarity 87.8%; Pred. No. 5.3e-73;

Matches 393; Conservative 0; Mismatches 41; Indels 12; Gaps 5;

QY 2134 AGTTACACGCAATTTGGAATTCACAGATACAGATGAGACGCTTAATCTCAGCTCTTG 2193
 DB 128 AGTTACACGCAATTTGGAATTCACAGATACAGATGAGACGCTTAATCTCAGCTCTTG 187
 QY 2194 CAGCTATTGGACCAAGAGAAAGACCACTAGAA-----TCGGAATTGAGGGCTTAA 2247
 DB 188 CAGCTATTGGACCAAGAGAAAGACCACTAGAAATTGCGCTTGGAAAGTAGGGCTTAA 247
 QY 2248 AAGACAACTCTTCTGCTTGGGACATCCAGCTGACAGCACCACCAACACTGCTATGTC 2307
 DB 248 AAGACAACTCTTCTGCTTGGGACATCCAGCTGACAGCACCACCAACACTGCTATGTC 307
 QY 2308 CAGGAATCAGGAATATGCTGCTCAGGACATTTGATTTTGTATGGAACAGGAACGGGAGA 2367
 DB 308 CAGGAATCAGGAATATGCTGCTCAGGACATTTGATTTTGTATGGAACAGGAACGGGAGA 367
 QY 2368 TGAAGTATTCGAGCTCTTATACCTGCGCTTCGGAAGTGAACACACACCTCCATCC 2427
 DB 368 TGAAGTATTCGAGCTCTTATACCTGCGCTTCGGAAGTGAACACACCTCCATCC 426
 QY 2428 ACATCTCTGCTATTACTGCGCAAGAACACAA--AGCATTTGTGCACTGCTGTAAT 2485
 DB 427 ACATCTCTGCTATTACTGCGCAAGAACACAAAGACATTAAGGACATTTGTGCACTGCTGTAAT 486
 QY 2486 TTCAATTTCTGGAATAATA-CACCAACATGAAGAGACATTTGTACGATTAGAA--CTTT 2542
 DB 487 TTCAATTTCTGGAATAATACTAATAGGAACACCATTTGTACGATTAGAACTTTT 546
 QY 2543 ATTAACTTACCTAT 2558
 DB 547 ATTAACTTACCTAT 562

RESULT 8
 B0220171 1410 bp mRNA linear EST 02-MAY-2002
 LOCUS B0220171

DEFINITION AGENCOURT_7571205 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6052107
5', mRNA sequence.
ACCESSION B0220171
VERSION B0220171.1 GI:20401571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1410)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M13307 row: a column: 04
High quality sequence start: 14
High quality sequence stop: 184.
Location/Qualifiers
1..1410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6052107"
/clone_lib="NIH-MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: PCMV-SPORT6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 346 a 423 c 330 g 307 t 4 others
ORIGIN
Query Match 11.3%; Score 288.6; DB: 14; Length 1410;
Best Local Similarity 84.9%; Pred. No. 1.8e-70;
Matches 376; Conservative 0; Mismatches 49; Indels 18; Gaps 4;
QY 1052 CCAACTGCTGCAGAGAGAGCAACAGTGGTGTGCTTCACTTCAAAACCA 1111
|||||
Db 1 CCAACTGCTGCAGAC--GGAGGAGACACAGCTGGAAGTGTGCTTCACTTCAAAACCA 58
QY 1112 CTTGTGACATCTGTGGCAACAGTGCACAGGTCTCAGTGCACAGTGAAGCCAGTT 1171
|||||
Db 59 CTTGTGACATCTGTGGCAACAGTGCACAGGTCTCAGTGCACAGTGAAGCCAGTT 118
QY 1172 GTCCTGGAAGAGAGAGTAACTGTCCTTGCAGAGTAACTTTGGAGAACTTCAGT 1231
|||||
Db 119 GTCTGTGGAAGAGAGTAACTGTCCTTGCAGAGTAACTTTGGAGAACTTCAGT 178
QY 1232 GCAGCTATTTGCTTCCATCTGTGAAACCTGTGTTCTTCTGCTGGAGCAACATCTG- 1290
|||||
Db 179 GCAGCTATTTGCTTCCATCTGTGAAACCTGTGTTCTTCTGCTGGAGCAACATCTG- 237
QY 1291 CAAGCTGTATTTGGAGTCCAGTTCAAATCAAACTTGGCCAGCGGCGCTGTCCCTTTC 1350
|||||
Db 238 CAAGCTGTATTTGGAGTCCAGTTCAAATCAAACTTGGCCAGCGGCGCTGTCCCTTTC 297
QY 1351 ACAACGAGCTGGAGTCCAGAGGAGTTCAGAGCAAGCAACTATTTCTATTTGTTTACGT 1410
|||||
Db 298 ACAACGAGCTGGAGTCCAGAGGAGTTCAGAGCAAGCAACTATTTCTATTTGTTTACGT 343
QY 1411 AGTTCAGAGCTTCAGAGGAGCAATGAAGAAAGAGTGCACCAATTTGACATTCCTCAAC 1470
|||||
Db 344 AGTTCAGAGCTTCCTGTGAGCAGCAGTGAAGAAAGCAGCAGCAGTTCATTTCCCAAC 403

QY 1471 ATTGACCAATTCAGAAATGTGAC 1493
|||||
Db 404 ATTGACCAATTCAGAAATGTGAC 426
RESULT 9
LOCUS BG213757 787 bp mRNA linear EST 21-APR-2001
DEFINITION R5733376 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG213757
VERSION BG213757.1 GI:13735444
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 787)
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,
E., Veloso,N., Kilka,A., Hess,J., Cothren,K., Lo,K., Offenbacher,
J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
JOURNAL 21227151
MEDLINE
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 443.
Location/Qualifiers
1..787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="H1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is H1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in H1080 under normal circumstances."
BASE COUNT 223 a 176 c 160 g 227 t 1 others
ORIGIN
Query Match 11.2%; Score 286.4; DB: 12; Length 787;
Best Local Similarity 97.5%; Pred. No. 4.9e-70;
Matches 312; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY 2241 GGGCTTAAGCAACCTCTTCTGTCGACATTCACCCGACAGCCACCAACAGTTG 2300
|||||
Db 4 GGGCTTAAGCAACCTCTTCTGTCGACATTCACCCGACAGCCACCAACAGTTG 63
QY 2301 CATGCTCAAAATCAGAGATCTGCTCAGAGGAGCTGATATTTGATGGAACAGGAA 2360
|||||
Db 64 CATGCTCAAAATCAGAGATCTGCTCAGAGGAGCTGATATTTGATGGAACAGGAA 123
QY 2361 CGGAGATGAAGTATTCGAGCTCTATACCTGGCCCTTCTGAAGTACCACTCCACTCT 2420
|||||
Db 124 CGGAGATGAAGTATTCGAGCTCTATACCTGGCCCTTCTGAAGTACCACTCCACTCT 183
QY 2421 TCATCCACATCTTGTATTTACTGCGCAAGAGACCAAGCATTTGCTGCTGCTCT 2480
|||||
Db 184 TCATCCACATCTTGTATTTACTGCGCAAGAGACCAAGCATTTGCTGCTGCTCT 243
QY 2481 GAAATTCATTTTGGAAATTA--CACCAATGAAGAGCATTTGTTAGATTAG--AA 2538
|||||
Db 244 GAAATTCATTTTGGAAATTAATCAACCAATGAAGAGCATTTGTTAGATTAG--AA 303
QY 2539 CTTTATTAACCTTACCTAT 2558

BC295865			
LOCUS	734 bp	mRNA	linear EST 21-FEB-2001
DEFINITION	602393337F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:4505179 5', mRNA sequence.		
ACCESSION	BC295865		
VERSION	BC295865.1	GI:13057927	

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 734)
TITLE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@ncl.nih.gov
Tissue Procurement: nih.gov
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM10378 Row: 1 Column: 20
High quality sequence stop: 721.
Location/Qualifiers
1. 734
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="4505179"
/clone_id="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 223 a 174 c 211 g 126 t
ORIGIN

Query Match 9.4%; Score 241; DB 12; Length 734;
Best Local Similarity 62.4%; Pred. No. 4.4e-57;
Matches 434; Conservative 0; Mismatches 245; Indels 16; Gaps 3;

QY 1730 AAACCTGATTTGGTGGACACATCATTCATGATGTAAGTAAACCATTTCTTTAT 1789
DB 21 AAACCTGATTTGGTGGACACATTCATGATGTAAGTAAACCATTTCTTTAT 1789
QY 1790 TGGAG-CTCTACAAAAGAACTTAGACATTTGTAAGCAATGACATTA 1848
DB 81 CGTGGCTTTGAGAGAGAGATGATGAGTAAAGCAAGCGATCAAGGACTGC 140
QY 1849 ACTCTGATGCTGTGACTTGTATCCCAAGCAACAGAGAGATGAGGCTTTAG 1908
DB 141 ATCCAGATGTAGTACGTCTCTATGCCACACAGCAAGGCTGAGATCTGTAG 200
QY 1909 AAAAAGTACGCAATTCGTCAGATGATGATGATGATGATGATGATGATGATGAT 1968
DB 201 AAAAAGTACGCAATTCGTCAGATGATGATGATGATGATGATGATGATGATGAT 260
QY 1969 TCCTGTAGTATACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2028
DB 261 AGCAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 320
QY 2029 AGCAGAGAAAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
DB 321 AGCAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 380
QY 2089 ATAAAGAGATTCAG 2148
DB 381 GACAG 440
QY 2149 AACTTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2208
DB 441 AGCTGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 500
QY 2209 GGAAG 2268

DB 501 GGAAG 560
QY 2269 GGAAG 2314
DB 561 GCGAG 620
QY 2315 CACGAG 2374
DB 621 CACGAG 679
QY 2375 TTCTGAGCTTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2409
DB 680 TTCACTGTGCTCTACAAAGCATTTCTTAAGTAAC 714

Search completed: February 17, 2003, 02:56:40
Job time : 3519 secs

This Page Blank (uspto)

XX Dikstein R, Yamit-heel A;
 XX WPI; 2000-256640/22.
 DR N-PSDB; AA290465.
 XX
 PT Polypeptide encoding TATA box binding protein associated factor II 105
 PT useful for treating e.g. cancers and inducing apoptosis has a dominant
 PT negative effect on the normal biological activity of the binding
 PT protein -
 XX
 PS Claim 7; Fig 2; 48pp; English.
 XX
 CC This represents a polypeptide comprising a (modified) fragment (I) of
 CC a TATA box-binding protein associated factor II 105 (TAFII105). A
 CC pharmaceutical composition comprising (I) or the polynucleotide or an
 CC inhibitor or antagonist of (I) is useful for treating cancers and
 CC inducing apoptosis in pathological cells. The composition is also useful
 CC for treating autoimmune diseases, inflammatory processes and viral or
 CC bacterial infections.
 CC
 XX Sequence 852 AA;
 SQ
 Query Match 99.9%; Score 4258; DB 21; Length 852;
 Best Local Similarity 99.9%; Pred. No. 9.2e-278;
 Matches 851; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 661 TRSOLKLEKLDLEKORRLEEREMILKAASRSNKDEPEQLRLKOKAKELQOLELAOI 720
 QY 721 QHRDANLTATAIGPRKKRPLESGIEGLKDNLASGTSITATKOLARRPITRICLDLI 780
 DB 721 QHRDANLTATAIGPRKKRPLESGIEGLKDNLASGTSITATKOLARRPITRICLDLI 780
 QY 781 FCMEOEREMKYSRALLYALLKZPLHSSIHILAIYCORRHALLHCEPISISGZHOHERA 840
 DB 781 FCMEOEREMKYSRALLYALLKZPLHSSIHILAIYCORRHALLHCEPISISGZHOHERA 840
 QY 841 LFTIRTLTLTY 852
 DB 841 LFTIRTLTLTY 852
 RESULT 2
 AAW31494
 ID AAW31494 standard; protein; 801 AA.
 XX
 AC AAW31494;
 XX
 DT 28-APR-1998 (first entry)
 XX
 DE Human hTAFII105 protein.
 XX
 KW TATA-binding protein associated binding factor 105; human; activator;
 KW hTAFII105; transcription factor; TFIID; transcriptional activation;
 KW antibodies; diagnosis; therapy; biopharmaceutical industry.
 XX
 OS Homo sapiens.
 XX
 PN US5710025-A.
 PD 20-JAN-1998.
 XX
 PF 02-OCT-1996; 96US-0725012.
 XX
 PR 02-OCT-1996; 96US-0725012.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Dikstein R, Tjian R;
 XX
 DR WPI; 1998-109818/10.
 DR N-PSDB; AA002872.
 XX
 PT DNA encoding human tata-binding protein associated factor - for
 PT producing recombinant protein
 XX
 PS Claim 1; Col 17-22; 12pp; English.
 XX
 CC This cDNA sequence represents a human tata-binding protein associated
 CC factor, hTAFII105, isolated from Daudi cell nuclear extracts. Tightly
 CC associated subunits (TAF's) are components of the transcription factor
 CC TFIID and are thought to mediate transcriptional activation. This encoded
 CC protein may be produced recombinantly from transformed host cells or
 CC purified from human cells. hTAFII105 specific binding agents such as
 CC specific antibodies could be used for diagnosis (e.g. genetic
 CC hybridisation screens for hTAFII105 transcripts), therapy (e.g. gene
 CC therapy to modulate hTAFII105 gene expression) and in the
 CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
 CC cell specific activators or other transcriptional regulators).
 XX
 SQ Sequence 801 AA;
 Query Match 93.6%; Score 3990; DB 19; Length 801;
 Best Local Similarity 99.9%; Pred. No. 9.1e-260;
 Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```

OY 421 VVFCWDHICKPVIGTIPVOIKLAQPCPVLSQAPGICPTGSSSKOLFSLFHVQOPSGGNEK 480
DB 422 VVFCWDHICKPVIGTIPVOIKLAQPCPVLSQAPGIPGSSSKOLFSLFHVQOPSGGNEK 480
OY 481 QVTTIHSSSTLTIOKCGOKTNPVTIIPTSPQPPASTIKOITLPKNKILSLQASPTQKNR 540
DB 481 QVTTIHSSSTLTIOKCGOKTNPVTIIPTSPQPPASTIKOITLPKNKILSLQASPTQKNR 540
OY 541 IKENVTSCFEDEDDINDVYTSMAGVNLNEENACILATNSELVGTLIOSCKDEPEFLGALQ 600
DB 541 IKENVTSCFEDEDDINDVYTSMAGVNLNEENACILATNSELVGTLIOSCKDEPEFLGALQ 600
OY 601 KRLLDYGKHDITELNSDAVNLSQATQERLCKLEKTRIAOHRMTTYVASENYIICSD 660
DB 601 KRLLDYGKHDITELNSDAVNLSQATQERLCKLEKTRIAOHRMTTYVASENYIICSD 660
OY 661 TRSOLKFLKLDLEKORNDLEERMLLKAASRSNKDEDEOLRLKOKAEILOOLEAQI 720
DB 661 TRSOLKFLKLDLEKORNDLEERMLLKAASRSNKDEDEOLRLKOKAEILOOLEAQI 720
OY 721 QHRDANLTATAIGPRKKRPLESGIEGLKNDNLASGTSSTLTATKOLHRRITRICLELDLI 780
DB 721 QHRDANLTATAIGPRKKRPLESGIEGLKNDNLASGTSSTLTATKOLHRRITRICLELDLI 780
OY 781 FCMEOEREMKYSRATYLLALK 801
DB 781 FCMEOEREMKYSRATYLLALK 801

RESULT 4
AAR56494
ID AAR56494 standard; Protein; 737 AA.
AC AAR56494;
DE 23-MAR-1995 (first entry)
XX TATA-binding protein-as-associated factor hTAFFI130.
KW TATA-binding protein associated factor; hTAFFI130; screening;
XX diagnostic; therapeutic; gene transcription regulation.
OS Homo sapiens.
XX WO9417087-A.
XX 04-AUG-1994.
XX 28-JAN-1994; 94WO-US01114.
XX 28-JAN-1993; 93US-0013412.
XX 30-JUN-1993; 93US-0087119.
XX (REGC ) UNIV CALIFORNIA.
XX
PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl KO;
XX
XX WPI; 1994-264019/32.
XX N-PSDB; AAQ070731.
XX
PT TATA-binding protein associated protein factors - and
PT corresponding nucleotide sequence and deriv. antibodies, useful
PT in screening, diagnostics and therapeutics
XX
PS Disclosure; Page 142; 180pp; English.
XX
XX The TATA-binding protein associated factor hTAFFI130 (including
XX specific antibodies and fusion products) are used in drug screening,
XX diagnostics and therapeutics. They are used in the development of
XX specific biochemical assays for screening compounds that agonise or
XX antagonise selected transcription factors involved in regulating

```

```

CC gene expression associated with human pathology.
XX
SQ Sequence 737 AA:
Query Match 30.7%; Score 1307.5; DB 15; Length 737;
Best Local Similarity 40.0%; Pred. No. 2,4e-79;
Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;

OY 13 PPRVSSG-----PRLPAPQIVAVKAPMTTITQFPANQLPPTGLIKNSGPTLM 61
DB 68 PTAFTSGIRATLPTVLAAPLPQ-----PQNPRTNQ-----NQLPQGMVLVSENGQL 118
OY 62 LVSPQQTWR-----AETTSNITSRPAPNDQVAKICTVPNSSQLIKKVAIPVKLL 115
DB 119 MI-POQALMOQAQAHQOPOTMAPRPAPTSAAPVOISTVQAAGPPIIAR-QVTP----- 172
OY 116 AQGTGVVTVTPRPSVQSAVPTSVTVTPGRPLNVT--TLKPSLASS-----TPS 168
DB 173 ----TTIIKOV---SQAQTVQPSATLQSPGVQPOLVLGGAQVMSLTGTAIVQGTQ 225
OY 169 NE-PNLKAENSAVAQVNLSPMLNENYKCKNFKLAMILKLACSGSGSPENMGQNKRIEOL 227
DB 226 RTVPGATTTSSADE-----TMENYKCKNFKLSTILKLASSGKSTETANKVELQNL 279
OY 228 LPAKTEAEETRKLYELKSSPOPHLVPLKKSVALROLPLPSQSFIOQCVOQ----TS 283
DB 280 LDKIEADEFTSRLYRELNSPPYLVPLKRSIPLALROLTPDSAFIOQSOQPPPTPS 339
OY 284 SDVVIATCTTGTTPVTTVSSQSEKSIIVSGATPAPTVSVQTLNPLAGPVGAKGV 343
DB 340 Q-----ATTALTAVLVSSVQRTACKTAATVTSALQPPVLSL----- 376
OY 344 VTLHVGPRPAATGTTAGTLLQTSKPLVTSVANTVTVSLQPEKPVSGTAVTSLPAY 403
DB 377 -----TQPTQGVGKQGOPTPLVIG-----QPPKP-----GALIRPQY 410
OY 404 TFGETSGAATCLPSVKPVYFPCWDHICKPVIGTIPVOIKLAQPCPVLSQAPGIPGSSSKO 463
DB 411 TLTQT-----PWVALRQPH-NRIMLTTPQOQL----- 437
OY 464 LFSLFHVQOPSGGNEKQVTTISHSSTLTIOKCGOKTNPVTIIPTSPQPPASTIKOITLP 523
DB 438 -----NPLQPVVVPVAVL 451
OY 524 PGNKIL---SLQASPTQKNRIKENVTSCFEDEDDINDVYTSMAGVNLNEENACILATNSEL 580
DB 452 PGTKALSAVSAQAQAAQKNKLEKPGGSGFRDDDDINDVASMAGVNLSEESARLATNSEL 511
OY 581 VGTLIOSCKDEPEFLGALQKRLIDYGKHDITELNSDAVNLSQATQERLCKLEKTRIAO 640
DB 512 VGTLTRSCKDEFTLQAPLQRLILEIGKKHGTTELHPDVVSYSVSHATQORLQNLVERISE 571
OY 641 IAOHRMTTYVASENYIICSDPTRSOLKFLKLDLEKORNDLEERMLLKAASRSNKDEP 700
DB 572 TAOQKNFSYKDDRRYDQASVYRQQLKFEQLDIEKORNDLEERMLLKAASRSNKDEP 631
OY 701 EQLRLKOKAKELQOLELAQIOHRDANLTATAIGPRKKRPLE-----SGIEGLKNDLNL 755
DB 632 EQLRLKOKAKEMQOELQORORDANLTATAIGPRKKRVDCPGSGSGEGSGESGVYP 621
OY 756 GTSSTLATQOLHRRITRICLELDIFCMEREMKYSRATYLLALK 801
DB 692 GSSGVGTPTROTTRIVNLRLIFCLENERETSHSLLYKFLK 737

RESULT 5
AAM06084
ID AAM06084 standard; Protein; 737 AA.
AC AAM06084;
XX
XX 27-JAN-1997 (first entry)
XX

```


PT Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs
 PS Claim 1; Column 111-116; 86pp; English.
 XX
 CC AA025018 represents TATA-binding protein associated factor (TAF)
 CC polypeptide, hTAF1100 (mol. weight 100kD). TAF peptides derived
 CC from hTAF1100 alpha, hTAF1130 beta, hTAF1140, hTAF1170, hTAF1100,
 CC hTAF1130, hTAF1250, hTAF1148 and hTAF1148 and nucleic acids encoding
 CC them, are used to modulate transcription, including transcription
 CC initiation. TAFs are nuclear proteins involved in RNA polymerase I,
 CC II and III transcription. The peptides act by binding to a different
 CC TAF, an activator, or TBP (TATA-binding protein) or competitively
 CC inhibiting association of a TAF domain with another compound, typically
 CC a protein like TBP or another TAF, an activator, or DNA.
 CC
 XX
 SQ Sequence 737 AA;
 Query Match 30 7%; Score 1307.5; DB 18; Length 737;
 Best Local Similarity 40.0%; Pred. No. 2,4e-79;
 Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;
 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTFQFPANLQLPQGVLIKNSGGLM 61
 68 PRTATSGIRATLPTVLAAPLPP-----PQNPNTIQ---NQLPFGMTLVSENGQL 118
 62 LVSPOQTWR-----AETTSNITSRPVAPNPQVICTVPSNSSQLIKVAVTPVKL 115
 119 MI-POALQMOQAQAHQAPQTMARPRAPTSPAPVQISTVQAPGRPIAR-QVTP----- 172
 116 AQGTIVTVTPKPSVQSAVPTSVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
 173 ----TTLIKOV---SQAQTVVQPSATLQKSPGVQPOLVIGGAQTAAGTATVQGTGPQ 225
 169 NE-PNKAENSAVQINLSPTMLENVKCKNFAMLIKACSGQSEPMQONVKYLEOL 227
 226 RTVPAGNTTSSATE-----TMENVKCKNFSTLIKAKLASSGKQSTETANKEVQNL 279
 228 LDKAIAEETKRLVYELKSSPOPHLVPLUKSVALRQLLPPNSQSFITQCVQO---TS 283
 280 LDKKIAEPTSRILYELNLSPPQYLPFLKRLPALRQLTPDSAAFITQSQOQPPPTS 339
 284 SDVAVTACTTPTVTPVTTVSSQSEKSIIVSGAAPTPTVQNTPLAGPVGAKGV 343
 340 Q-----ATTALPAVLLSSVORTAKTATVTSALQPPVLSL----- 376
 344 VTLHVGPTAATGTTAGTGLTQTSKPLTVSVANTVTVTSLOPEKPYVSGTAVLSLPAV 403
 377 -----TQPTQVGVKQGGPPPLVITQ-----QPPKP-----GALLIRPPV 410
 404 TBEETSGAALCLPSPVPSFCMDHICKPYTIGTPVOIKLAQPGPVLSOPAGIPTGSSSKO 463
 411 TLNQT-----PMVALRQPH-NRIMLTTPQOQL----- 437
 464 LFSLFHVQOPSGGNEKQVTTISHSSTLTQKGGKTMPTNTIIPTSQPPASILKQITL 523
 438 -----NPLQPPVPAVLA 451
 524 PGKNTL---SLQASPQOKNRIKENVTSCFREDDDINDVTSMAGVNLNEENACILATNSEL 580
 452 PGKALSAVSAQAAMAKKKEKPEGGSPFRDDDDINDVAMAGVNLSEEARILATNSEL 511
 581 VGLTIOKSCDEPFLFGALQKRLIDIGKKKIDITELNSDAVNLISQATQERLKGLEKLA 640
 512 VGLTIOKSCDEPFLFGALQKRLIDIGKKKIDITELNSDAVNLISQATQERLKGLEKLA 640
 641 IAOHRTTYKASENYILCSPTRSQLEKLDLEKOKRLEKREMLKAKRSKNEDEP 700
 572 TAOQKPFSTKDDRYDQASVRAQLKFFEQLDIEROKRDEKREILMRAKRSRQEDP 631
 701 EQLRLKOKAKKELQOLELAQIQHRDANLTATAAGPRKKRPLE-----SGTEGKDNLLAS 755

DB 632 EQLRLKOKAKKEMOQOELQMRGRDANLTATAAGPRKKRVDCPSSGAGSGSPSV 691
 QY 756 GTSSLTATFKQLHRPRITRICLRDLIFCMEQERKKYSRALYALIAIK 801
 DB 692 GSSGVGTPTQFTRITRIVNLRDLIFCLENDRETSLSLLYKAPIAK 737
 RESULT 7
 ID AA082954 standard; Protein; 1023 AA.
 XX
 AC AA082954;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human homologue of Mpr1 protein target for antifungal compound.
 XX
 KM Antifungal; fungal gene transcription; RPC34; POB3; TFA2; NAB2;
 KM MPT1; MPR2; BOS1; POL30; RSB2; SQT1; MRM1; TRB1; SPC98; BPR2; RNL1;
 KM GCD7; SK16; NIP1; LCP5; NCE103; ECO1; ORC2; CNS1; YPD1; TIM10; SRB4;
 KM Yeast; fungus.
 XX
 OS Homo sapiens.
 XX
 PN WO200202055-A2.
 XX
 PD 10-JAN-2002;
 XX
 PF 28-JUN-2001; 2001WO-US20592.
 XX
 PR 29-JUN-2000; 2000US-215164P.
 PR 10-AUG-2000; 2000US-224457P.
 XX
 PA (ANAD-) ANADYS PHARM INC.
 XX
 PI Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;
 PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long F;
 PI Davidov E, Thompson CM;
 DR N-PSDB; ABK32842.
 DR WPI; 2002-147962/19.
 XX
 PT Screening candidate antifungal compound for interaction with essential
 PT protein, modulation of essential protein activity, binding to essential
 PT protein, by contacting protein with test compound and determining
 PT effects
 PS Claim 1; Figure 79; 522pp; English.
 XX
 CC The invention describes a method of screening a candidate antifungal
 CC compound for interaction with essential proteins (EP) or for modulation
 CC of EP activity e.g fungal gene transcription. The proteins tested in the
 CC invention include RPC34, POB3, TFA2, NAB2, MPT1, MPR2, BOS1, POL30, RSB2,
 CC SQT1, MRM1, TRB1, SPC98, BPR2, RNL1, GCD7, SK16, NIP1, LCP5, NCE103,
 CC ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans
 CC and human homologues. The method involves contacting a culture with one
 CC or more test compounds and determining the effects on the growth or
 CC viability of the culture of cells which preferably comprises fungal cells
 CC or yeast cells. Preferably the identified compounds interact with, or
 CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor
 CC compounds identified by the method are useful for preventing or
 CC inhibiting fungal, particularly C. albicans growth in culture or in a
 CC mammal. The antifungal agents interact with essential fungal elements
 CC that can be used to treat fungal infection by preventing the growth and
 CC preferentially killing the fungi, but does not inhibit the biological
 CC activity of mammalian homologues. This amino acid sequence represents a
 CC target protein used to test the antifungal compounds, described in the
 CC method of the invention.
 XX
 SQ Sequence 1023 AA;
 Query Match 27 1%; Score 1157.5; DB 23; Length 1023;
 Best Local Similarity 37.6%; Pred. No. 4.8e-69;

Matches 310; Conservative 97; Mismatches 168; Indels 249; Gaps 26;

```

QY 13 PPKVSG-----PRLPAPQIVAKPNTTITQFPANLQLPQGVLIKNSGPI 61
D 414 PTATSGIRATLPPLVLRPQ-----PQNPTNIG---NFQLPQGVLRSENGQL 464
QY 62 LVSPQGVTR-----AETSNITSRAPVANPOTVKTICVTPNSSQLIKVAVTPVK 115
D 465 MI-PQALAMQQAQAHQAPOTTMARPRATPSAPPVQISTVQABGTPIIAR-QVTP----- 518
QY 116 AQTGTVTTPKPSVQSVAVPTSVTPGKPLNTVT---TLKPSISGAS-----TPS 168
D 519 -----TTIKOV---SQOATVQPSATLORSPPVQPOLVGAADPTASLGTAATVQGTGP 571
QY 169 NE-PNLKENSAAVOINLSPMLNKKCNKFLAMLIKLACSGSGSPMGQNKIYEOL 227
D 572 RTVGATTTSSAATE-----TMENVKCKKFLSTLIKLSGSGSTTANVEIYNL 625
QY 228 LDAKTEBEETRKLYVELKSPQPHLYPFLKKSVALRQLLPNSQSFIQOCVQOTSSDMV 287
D 626 LDGTEADEFTSRLRYREINSSPQYLVFLKRSIPALROLPPDSAAFIQSQOQPPPP-- 683
QY 288 IATCTVTTSPTVTTVSSQSEKSIISGATAPRTVSVOTLNPAGPVAKKGVTLH 347
D 684 -----TSQATTA--LTAVALSS-----SVQ-----R 702
QY 348 SVGPATATGGTTAGTGLQTSKPLVSVANTVTVSLOPEKPVSGVATLSTLPAVTFGE 407
D 703 TAGKTAAT-----VTSALQP--PVLSLT----- 723
QY 408 TSGAALCLPVKPVVSCWHDICKPVIGTIPVQI---KLQAGPVLSQPAIGTPGSSSKOL 464
D 724 -----TPQVGVGKOGOPTPL----- 739
QY 465 FSLFHVAVQPSGNEKQVTTISHSSTLIQKCGKTPMVTNIIPTSPQPPASILK--QIT 522
D 740 -----VIOQPP-----KPCALIRPPQVT 757
QY 523 LPGNKILSLQASPTQKRIKENVTSCFEDDINDVTSMAGVNNENACILATNSELVG 582
D 758 LTQTPVALR---QPHNRIM-----LTPQOVNLSEESARILATNSELVG 799
QY 583 TLVQSCDEPFLFQALQKRLIDGKHDITELNSDANVLSQATQOEKRLKELKRAIA 642
D 800 TLVRSCEDEFLLQAPLQRLLEIGKKGITELHPDVVSVYSHAQOQLQVLEKISETA 859
QY 643 QHRMTYKASENYILCSDTRSQLEKLELDQLEKORQDLEERMLTKAKRSNKDEPQ 702
D 860 QQKNESTKDDRYEQASDVRAQLKFFEQDQLEKORQDQERETILMRKRSRQEDPEQ 919
QY 703 LRLKQAKELQOLEIAQIOHRDANLTATAIGPRKKRPLE-----SGIEGLKMLLASGT 757
D 920 LRLKQAKEMQOQOLAQRQDANLTALAIIGPRKKRQVDPGPGSGEGSGPSVPGS 979
QY 758 SSLTATQQLHRRITRITCLDLIFCMQDEREMKSRALYLALAK 801
D 980 SGVGTPTQRTQRTIRVNLRLIFCLENERETSHSLLYKAFK 1023

```

RESULT 8

AA56487 standard; protein; 921 AA.

AA56487:

23-MAR-1995 (first entry)

TATA-binding protein-associated factor dTAFFII110.

TATA-binding protein associated factor; dTAFFII110; screening; diagnostic; therapeutic; gene transcription regulation.

Drosophila.

```

XX XX W09417087-A.
XX XX
XX XX 04-AUG-1994..
XX XX
XX XX 28-JAN-1994; 94WO-US01114.
XX XX
XX XX 28-JAN-1993; 93US-0013412.
XX XX 30-JUN-1993; 93US-0087119.
XX XX
XX XX (REGC ) UNIV CALIFORNIA.
XX XX
XX XX Comal L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
XX XX Tjian R, Wang E, Weinzierl ROJ;
XX XX WPI: 1994-264019/32.
XX XX N-PSDB: AAQ0724.
XX XX
XX XX TATA-binding protein associated protein factors - and
XX XX corresponding nucleotide sequence and deriv. antibodies, useful
XX XX in screening, diagnostics and therapeutics
XX XX
XX XX Disclosure; Page 56-61; 180pp; English.
XX XX
XX XX The TATA-binding protein associated factor dTAFFII110 (including
XX XX specific antibodies and fusion products) are used in drug screening,
XX XX diagnostics and therapeutics. They are used in the development of
XX XX specific biochemical assays for screening compounds that agonise or
XX XX antagonise selected transcription factors involved in regulating
XX XX gene expression associated with human pathology.
XX XX
XX XX Sequence 921 AA;

```

Query Match 18.2%; Score 775; DB 15; Length 921;

Best Local Similarity 29.6%; Pred. No. 2.2e-43; Matches 258; Conservative 126; Mismatches 308; Indels 178; Gaps 26;

```

QY 25 PQIYAVKAPNTTITQFPANLQLPQGVLIKNSGPIML-----VSPQOVTVAETTS 76
D 133 PQSPITSLTINTGQTPA-----LLKTDNGFOLLRVGTTGPPIVQTITTSNNS 184
QY 77 NITSRPANPQVYKICTVNSSQ-----LKKVAVTPYKLAQIGTVVTVTP 127
D 185 NITSRNPPTTO--IRLOTPAASMTNTATSNIVNSVSGVANSQPPHILQOLNAQ 243
QY 128 KPSSVQSAVPTSVYTVPGKPLNTVTLKPSLGASSTPENRPLKENSAAVQINSP 187
D 244 APQLPQITQIOTIIPAQSQOQOVNNSAGGATRAVSTTA-----ATT 287
QY 188 TMLNFK-KCKNFLAMLIKLACSGSGSPMGQNKIYEOLLDKIEAEFTRKLYVELK 246
D 288 TQGGTKKCKKFLANLEL---STREPRKVENRTLQELVNNVDEEECDRLERLNL 345
QY 247 SSPQHLVPLKKSVALRQI-----LPSQSFIQ--- 276
D 346 ASPQCLGLFKSLPLRLQALYKELYIEGIKPPQHVGLAGLSQOLPKIQAIQIRIG 405
QY 277 -----QCVQOTSSDMVIACTTIVTTSPTVTTVSSQSEKSIISGATAPRTVS---VQT 329
D 406 PSQTTTIGQTOVRMI--TPNALGTPRTTIGHTTISKQPPN--IRLPTAPRLVMTGIRT 460
QY 330 LNPAGPVGAKAGVTLHSGPPTAATGTTAGTGLQTSKPLVTSVANTVTVSLOPEK 389
D 461 QIP-SIQVPGQANIVQIR--GQHAQLORTSGVQIRATTRP-----PNSVPTAN----- 506
QY 390 VVSGTAATVLSLPAYTFGTSQAALCLPSVKPVVSCWHDICKPVYIGPVQIKLAQPGV 449
D 507 -----KLTAVKVGQTIKAI-TPSLHP-----PSLAISGCP-----PPTPTL 543
QY 450 SQPAGIPGSSSK---QFSLFHVAVQPSGNEKQVTTISHS----- 489
D 544 SVLSTLNASATTTLPISLPVTHLPPEALRARQMONLHNSHMFDAKVEIKAPSLHP 603

```


QY 490 -----TLTIKCGOKTNPVNTIPTSOPPPASILKQITLPKNTLSIQ-----AS 534
 Db 604 PHMERINASTLPFGAKTM-----ARPPAINKAIGKKRDKADAMEMAKLNTSSGGA 654
 QY 535 PTOKNRKIKENVTSCEFNDEDINDVTSAGVNLNEENACILATNSELVGTILQSCDEPFL 594
 Db 655 SAANSFEOQSSMSMYGDDDDINDVAMGCVNLAEESORILGC-TENIGTOIRSCKEVEFL 713
 QY 595 FIGALOKRILDIKKHIDITELNSDANVLISQATQERLGLLEKTLVIAOHRMTTYKASEN 654
 Db 714 NLPSTLOARIRATISEAGLEPDSQDVAVLISHACQERLKNVETKLVIAEHRIDIVIKIDPR 773
 QY 655 YILCSPTRSQLEKLELDLEKORLEEREMTLKAASKNKEDPEQOLKOKAKELOQ 714
 Db 774 YEPADVROQIKFLELDKAEQKRHELEEREMTLRAKSSRSRVEDPEQAKMKARAKEMOR 833
 QY 715 LELAIOHRDANLTATATAGPRKKRPLE-----SGIEGLKDNILASGTSLSLATKOLHRP 769
 Db 834 AEMEELRQDANLTALQALGPRKKLKLDEGTYSVGAGSSGGCVLSSSGSAPPTL-----RP 889
 QY 770 RITRICLRDLIFCMEQEREMKYSRALYLALLK 801
 Db 890 RIKRVNLRLDMLFYMEOERECSSMLFKTYLK 921

RESULT 9

AAW06077 standard; Protein; 921 AA.

AAW06077;

27-JAN-1997 (first entry)

Drosophila TATA-binding protein associated factor dTAFII110 protein.

Drosophila; TATA-binding protein; TBP associated factor; TFIID;

RNA polymerase II; transcription; messenger RNA; nuclear fraction;

holoenzyme; lambda-gt11; expression library.

Drosophila melanogaster.

US534410-A.

09-JUL-1996.

28-JAN-1993; 93US-0013412.

28-JAN-1994; 94US-0188582.

28-JAN-1993; 93US-0013412.

30-JUN-1993; 93US-0067119.

(RBCG) UNIV CALIFORNIA.

Comal L, Dynlacht BD, Hoey T, Rupert S, Tanese N;

Tjlan R, Wang E, Weinzierl ROJ;

WPI: 1996-333245/33.

N-PSDB: AAT42210.

Screen for cpds. that bind human TATA-binding protein associated

factor - by testing ability to bind to polypeptide fragments of the

factor, useful as (ant)agonists of transcription factors involved in

disease.

Examples; Column 27-36; 86pp; English.

This is the amino acid sequence of the Drosophila TATA-binding protein

(TBP) associated factor (TAF) designated TAFII110. The protein is a

component of the TFIID fraction required for reconstituting RNA

polymerase II in vitro transcription activity. The encoded protein

has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.

based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.

The invention relates to purified proteins involved in transcription

CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIB, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.

SQ Sequence 921 AA;

Query Match 18.2%; Score 775; DB 17; Length 921;

Best Local Similarity 29.6%; Pred. No. 2,2e+43; Mismatches 178; Gaps 26;

Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

QY 25 PQLVAVKAPNTTIIOPFANQLPPGVILIKNSGPMPL-----VSPQGVTRAEETS 76
 Db 133 PQSPSTLSLNTGQTPA-----LVYTDNGFOLLKRGTTGTPPTVQITTNSSNS 184
 QY 77 NITSRPVAPNPOTVKICIVPNSSSQ-----LKKVAVTPVKLAQIGTVVTVTP 127
 Db 185 NTSSTNHPTTQ-IRLOQVPAASMTNTATSNIIYNSVASSGVANSSQPPHLTQLMAQ 243
 QY 128 KPSSVQSVAPTSVTVTPCKPLNTVYTLKPPSSIGASTPSNEPNLKAENSAVQNLSP 187
 Db 244 APQLPOTQIOTIPADQSOQOQVNVNVSAGCTAVAVSTTA-----ATP 287
 QY 188 TLEENK-KCKNFLAMLIKACSGSOSPEMGQVKKLVEDLAKIEAEFTKLYELK 246
 Db 288 TQGNTRKEKCKRFLANIEL-STREKPYEKVNRVILQELVANNVPEFCRLERLN 345
 QY 247 SSPQPLVPLPKRSVVALROL-----LPSQSFTQ-- 276
 Db 346 ASPQPCIGELKRLSLPLRALYKELVIGIKPPQHVHGLAGLSQQLPKIQAOIRPIG 405
 QY 277 -----QCVQOQSSMNVATCTTWTTSVTVTVTVSSQSEKSIIVSGATARTVS---VOT 329
 Db 406 PGGTTTIGQVQVMT-TPNALGTPTTIGHTTTSKOPN---IRLPARLNTGQIR 460
 QY 330 LNPAGPVGAKGVNVLHSGPTAAGTAGTGLLQTSKPLVTSVANTVTVSLQPEKP 389
 Db 461 QIP-SLOVPGQANIVQIR-GPQHAQOQRGTSQVQIRATTPR-----PNSVPTAN----- 506
 QY 390 VVSGTAVTSLPRAVTGEMTSGAALCLPSVKPVVVSFCMDHICKPVIGTPVOIKLADQEPVL 449
 Db 507 -----KLTAIVKVGQDQIKAI-TPSLHP-----PSLAISGSP-----PPTPL 543
 QY 450 SOPAGIPTGSSSK---QLPSLPHVVOQPSGCKNEKQVTTISHS----- 489
 Db 544 SVLSTLNSASTTLPPLPSLFTVHLPPALRARQMONSLNHNHNDAKLVEIKADSLHP 603
 QY 490 -----TLTIKCGOKTNPVNTIPTSOPPPASILKQITLPKNTLSIQ-----AS 534
 Db 604 PHMERINASTLPFGAKTM-----ARPPAINKAIGKKRDKADAMEMAKLNTSSGGA 654
 QY 535 PTOKNRKIKENVTSCEFNDEDINDVTSAGVNLNEENACILATNSELVGTILQSCDEPFL 594
 Db 655 SAANSFEOQSSMSMYGDDDDINDVAMGCVNLAEESORILGC-TENIGTOIRSCKEVEFL 713
 QY 595 FIGALOKRILDIKKHIDITELNSDANVLISQATQERLGLLEKTLVIAOHRMTTYKASEN 654
 Db 714 NLPSTLOARIRATISEAGLEPDSQDVAVLISHACQERLKNVETKLVIAEHRIDIVIKIDPR 773
 QY 655 YILCSPTRSQLEKLELDLEKORLEEREMTLKAASKNKEDPEQOLKOKAKELOQ 714
 Db 774 YEPADVROQIKFLELDKAEQKRHELEEREMTLRAKSSRSRVEDPEQAKMKARAKEMOR 833
 QY 715 LELAIOHRDANLTATATAGPRKKRPLE-----SGIEGLKDNILASGTSLSLATKOLHRP 769
 Db 834 AEMEELRQDANLTALQALGPRKKLKLDEGTYSVGAGSSGGCVLSSSGSAPPTL-----RP 889
 QY 770 RITRICLRDLIFCMEQEREMKYSRALYLALLK 801

Db 890 RIKRVNLDMLFYMEQERFCRSMLEFITYLK 921

RESULT 10
AAM25028
ID AAM25028 standard; Protein: 921 AA.

XX
AC AAM25028;
XX
DT 08-OCT-1997 (first entry)
XX
DE TATA-binding protein associated factor, dTAFl110.
XX
XX TATA-binding protein associated factor; TAF; nuclear protein;
XX RNA polymerase transcription; TATA-binding protein; TBP;
XX Initiation.
XX
OS Drosophila sp.
XX
PN US5637686-A.
XX
PD 10-JUN-1997.
XX
PE 28-JAN-1993; 93US-0013412.
XX
PR 28-JAN-1994; 94US-0188582.
XX 28-JAN-1993; 93US-0013412.
PR 30-JUN-1993; 93US-0087119.
PR 09-MAY-1996; 96US-0646715.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
DR WPI: 1997-319113/29.
XX
XX N-PSDB: AAT/9604.
XX
XX Nucleic acids encoding human TATA-binding protein associated factor
XX (TAF) peptide(s) - for production of recombinant peptide(s), used
XX for modulating transcription of TAFs
XX
XX Example 1: Column 35-40; 86bp; English.
XX
XX AAM25028 represents TATA-binding protein associated factor (TAF)
XX polypeptide, dTAFl110 (mol. weight 110kD). TAF peptides derived
XX from dTAFl110 alpha, dTAFl110 beta, dTAFl140, dTAFl160, dTAFl180,
XX dTAFl110, dTAFl1150, and dTAFl1250, their human equivalents and
XX nucleic acids encoding them, are used to modulate transcription,
XX including transcription initiation. TAFs are nuclear proteins involved
XX in RNA polymerase I, II and III transcription. The peptides act by
XX binding to a different TAF, an activator, or TBP (TATA-binding protein)
XX or competitively inhibiting association of a TAF domain with another
XX compound, typically a protein like TBP or another TAF, an activator,
XX or DNA.
XX
XX
XX
SQ Sequence 921 AA;

Query Match 18.28; Score 775; DB 18; Length 921;
Best Local Similarity 29.68; Pred. No. 2.2e-43;
Matches 256; Conservative 128; Mismatches 308; Indels 178; Gaps 26;

QY 25 POIYAVKAPNTTITOPANLQLPGTVLIRKNSGPLM-----VSPQOTVRAETTS 76
DB 133 PQSFTSLSTLNTGCTA-----LLVKTNDGFOILRVCTTGPPIVOTITINTSINS 184
QY 77 NITSRAPVAPNPQTVKICTYVNSSQ-----LKKVAVTPYKKLAQIGTIVTVTP 127
DB 185 NTSSTNHPPTTQ-IRLQTVPAASMTWTATSMIYNSVASGVSANSGOPHILQLNAQ 243
QY 128 KPSSVQAVAVPTSVVYTPGKPLNTVTLTKRSSSLGASSTPNERIKAKENSAVAQINISP 187
DB 244 APQLPQITQITIPAOOSQOQVNNVSSAGGTATAVSSTTA-----ATT 287

QY 188 TMLENVK-KCKNFLAMLIKILACSGSQSPDMGQNKVLEQLLDKIEAEFEFRKLYVELK 246
DB 288 TOGNTKCKCKRFLANLIEL--STREPKPVKENVNTLQELVYNNVNEFEEDCRLERLLN 345
QY 247 SSPQHLVPELKKSVVALROL-----LPSQSFQ--- 276
DB 346 ASPQCLIGFLKKSLPILRLQALYKEVIEGIRKPPQHLGLAGISQOLPRIOAQIRIG 405
QY 277 ----QCVOQTSDDWAIACCTTIVTTSPPVTTTVSSQSEKSIYSGANAPRTVS---VQT 329
DB 406 PSQTTTIGQTVRMT--TPNALGTPRPTIGHITTSKQPN---IRLPAPRLVNTGIGRT 460
QY 330 LNPAGPYGAAVYVTLHSGVPTAATGCTAGTGLQTSKPLVTSVANTVTVSSLOPEKP 389
DB 461 QLP-SLQVPGGANIVQIR--GPNHQLORTGVSQVRAITRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPVAVTGGTSGAALCPBVKPVVSVFCMDHICKPVIGTPVQIKLAPGVL 449
DB 507 -----KLTAVKVGQTIKAI-TPSLHP-----PSLAATSGGP-----PPTPL 543
QY 450 SQPAGIPRTGSSSK---QLFSLFHVYQPSGNGEKQVTTISHS----- 489
DB 544 SVLSTLNSASTTTLPIPSLPVTHLPEALRAREQKQNSLNNHSHFDKALVEIKAPSLHP 603
QY 490 -----TLTIQKCGQKTPVNTIIPTSOPPASILKQITLPGNKILSIQ-----AS 534
DB 604 PHMERINSLRPIGAKTM-----ARPPAIKAIKAKKKRDMEMDAKNTSSGGAA 654
QY 535 PTQKNRIKENVTSCEFDEDINDVTSMAGVNLNEENACILATNSELVGLTIQSCDEPFL 594
DB 655 SAANSFFQSSMSMWGDDINDVAAAGVNLAEESQRIIC-TEINIGTQIRSCDEVEL 713
QY 595 FIGAQRIIDIGKKHDTIELNSDAVNLISQATQRLGLEKLTALAQHRTTYKAEN 654
DB 714 NLPSTQARIRAITSEAGDEPSQDAVLISHACERLNIYEKLVAVIEHRIDVIKIDPR 773
QY 655 YILSDTRSQKLFLEKLOLEKQKDEEREMILKAASRSNKEDPEQLRKOKAKEIQ 714
DB 774 YEPKADVGOIKFLFELDKAQKREHELEREMILRAASRSRVEDPEQAKKAKKEKOR 833
QY 715 LELAOIQRDANLATAAIGPKRRKPLE-----SGIEGLKDNLASGTSSTLATQKLRP 769
DB 834 AEMELRQRDANLFLAQAIGPKRKLKIDGETVSSGAGSGGVLSSGSAPPTL---RP 889
QY 770 RITRCLDLFLCMQEREMKSRALYLALAL 801
DB 890 RIKRVNLDMLFYMEQERFCRSMLEFITYLK 921

RESULT 11
ABB61528
ID ABB61528 standard; Protein: 921 AA.

XX
AC ABB61528;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11376.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.

Db 2227 QVO---TTTSQPI---PIPHMSLOIPISOGQPOPOVOSSTQTLs-----SGQTLN 2272
 QY 481 QVTTTSHS-STLTQKCGKTMPTVNTIIPTSQF-PPASILKQITLPGNKILSLQASPTQK 538
 Db 2273 QVSVSSPSRPQLOIQ-----POPOVIAPVPOLOQOVVLSQI--OSQVVAQIOA---OQ 2321
 QY 539 NRIKENV-----TSCFRDEDDINDVTSMAGVNLNEENACILATNSLVTGLIOSCK 589
 Db 2322 SGVPQIKIQLPIQIOSSAVQTHQLONVVTVQAAVQEDL-----QRVQQLR 2369
 QY 590 DEPFLEIGALQKR---ILDIGKKHDTTELNSDAVNLIISOATQERLGLLEKLTALIAQH-- 644
 Db 2370 DQO-----QKKQOQOIEIKRHTLQASNQ-----SELQKQ---VVKHNAVIEHLK 2413
 QY 645 ---RMTTYKASEN--YILCSDFRSQKLEKLDQLEKQKRDLEEREMLLKAASRSNKED 699
 Db 2414 QKSMTPAREERQRMIVCNQVWKYI--LDKIDKEKQ-----AAKKRRRES 2459
 QY 700 PRQLRKQKAKELQOL 715
 Db 2460 VEOKRSKQNAKTKLSAL 2475

RESULT 14

AA57452
 ID AA57452 standard; Protein; 2907 AA.

AC AA57452;

DT 22-FEB-2000 (first entry)

DE Human transcriptional regulatory factor SEQ ID NO:1.

KM Human; transcriptional regulatory factor; TCOAL; BLAST detection;

KW bromo-domain; cell proliferation; cancer.

OS Homo sapiens.

PN W09957143-A1.

PD 11-NOV-1999.

PE 30-APR-1999; 99WMO-JP02340.

PR 30-APR-1998; 98JP-0137631.

PA (CHUGAI) CHUGAI RES INST MOLECULAR MEDICINE INC.

PI Jones MH;

DR WPI: 2000-052940/04.

DR N-PSDB: AA239032.

PT Transcriptional regulatory factor containing a bromo domain and gene

TCOAL encoding it

PS Claim 1; Page 48-68; 154pp; Japanese.

CC The present sequence represents a human transcriptional regulatory factor

containing a bromo domain. The factor interacts with proteins involved

in the chromatin-mediated transcription regulatory mechanism. It binds

to hSNF2h, hSNF2L and NCOA-62/Skip. It can be used for screening

compounds binding to it and acting as agonists or antagonists, which

are potentially useful for the treatment and prevention of cancer and

other cell proliferation disorders.

XX Sequence 2907 AA;

Query Match 6.0%; Score 255.5; DB 21; Length 2907;

Best Local Similarity 22.8%; Pred. No. 1e-07;

Matches 181; Conservative 101; Mismatches 298; Indels 213; Gaps 37;

QY 7 VAPVSAIPRVSSGPRLPAPQIYAVAKAPNTTITQIPANQLPCTVLI--KNSNGPLMLYS 64

Db 1938 VAPISG--SYTGTCKMVL--TTVGSFATVTEQONKNEHOTFAWNAQGSNSG---VVO 1990
 QY 65 PQQVTRAETTSNITSRAVPA--NPQTVKICTVNP-----SSQLIKKVVATPVKTL 115
 Db 1991 VQCKVLGIIPSSGTISQGTTSFQPRATVITIRPNSSGSGTSSNSGVITGPQIRP----- 2046
 QY 116 AQITGVVTVTPKPSVQSAVAPTSVTVTPGKPLNTVTTL---KP--SSLGASSTPSNE 170
 Db 2047 ---GNTVIRT-PLQGSFLGKAIITRPVMVQPGAPQVMTQIIRGQPVSTVASPNTVST 2102
 QY 171 PNKAENSAVQINLSPTMLENKKCNFLAMLKILACSSGSSQPEMGNKKVLEQLDA 230
 Db 2103 PGOKSLTSA-----TSSNTIOS-----SASQPPRPOQGVKLTMAQLTQ 2141
 QY 231 KIEAEETRLKYLELKSSPPHLPFLKSSVALROLPLPSQSEFI---QCVQO----- 281
 Db 2142 LTQGHGNGQGLVYVIGQCGQ-----TTQQLDLIPGVIVLPEPGQQLMAQAMPNG 2191
 QY 282 TSSDMVIANCTTCTVTVTPSPVTVTVSSS-----QSEKSI--TVSGATA 321
 Db 2192 TVQRFLETPLATATATASTTTTVSTTAAGTGEORQSKLSPOQVHQDKTLPPAQSSSVG 2251
 QY 322 PRTVSVQTLNPLAGP-----VGAKAGVYTLHSVG-----PTAATGTTA 360
 Db 2252 PAKAQPTQAPSAKPOPOQPOSPAPQPEVOTQPEVOTQTVSSHVSEAPTHA----- 2305
 QY 361 GTGLQTSKPLVTSVANTVTVTSLOPEKPYSGTAATLSLPAYTFGETSGAATCLPSVKP 420
 Db 2306 ---QSKRPVAAQS-----QPSNVQGSQPVAVQSPQTRIRPSRISQLSQCGQS 2352
 QY 421 VSEFCWDHICKPVYIGTPVQIKLAQPGFVLSQAPAGIPGSSSKOLFSLFHVQOPSGCNEK 480
 Db 2353 QVO---TTTSQPI---PIPHMSLOIPISOGQPOPOVOSSTQTLs-----SGQTLN 2398
 QY 481 QVTTTSHS-STLTQKCGKTMPTVNTIIPTSQF-PPASILKQITLPGNKILSLQASPTQK 538
 Db 2399 QVSVSSPSRPQLOIQ-----POPOVIAPVPOLOQOVVLSQI--OSQVVAQIOA---OQ 2447
 QY 539 NRIKENV-----TSCFRDEDDINDVTSMAGVNLNEENACILATNSLVTGLIOSCK 589
 Db 2448 SGVPQIKIQLPIQIOSSAVQTHQLONVVTVQAAVQEDL-----QRVQQLR 2495
 QY 590 DEPFLEIGALQKRILDIGKKHDTTELNSDAVNLIISOATQERLGLLEKLTALIAQH----- 644
 Db 2496 DQO-----QK-----KKQOQIEINVTNPSSKLLIKVEIIOQVYVWKHNAVIEHLKQK 2542
 QY 645 RMTTYKASEN--YILCSDFRSQKLEKLDQLEKQKRDLEEREMLLKAASRSNKEDPEQ 702
 Db 2543 SMTPAERENQRMIVCNQVWKYI--LDKIDKEKQ-----AAKKRRRESVEQ 2588
 QY 703 LRLKQKAKELQOL 715
 Db 2589 KRSKQNAKTKLSAL 2601

RESULT 15

AA57141
 ID AA57141 standard; Protein; 2035 AA.

AC AA57141;

DT 19-MAR-1995 (first entry)

DE Host cell factor protein.

KW Herpes simplex virus; herpes virus; VP16; immediate early gene;

host cell factor; virus infection therapy; cellular protein;

OS Homo sapiens.

QY key Location/Qualifiers

FT Active-site 21..31
FT /note- "peptide R60"
FT 168..186
FT /note- "peptide R37"
FT 333..340
FT /note- "peptide R52"
FT 426..449
FT /note- "peptide 362"
FT 511..526
FT /note- "peptide 329"
FT 578..594
FT /note- "peptide 223 first sequence"
FT 594..611
FT /note- "peptide R26 1st peptide"
FT 611..623
FT /note- "peptide 223 2nd sequence"
FT 723..731
FT /note- "peptide 318"
FT 802..813
FT /note- "peptide 299"
FT 813..820
FT /note- "peptide 268"
FT 836..847
FT /note- "peptide R26 2nd sequence"
FT 1010..1031
FT /note- "THE TNT repeat 1"
FT 1072..1093
FT /note- "THE TNT repeat 2"
FT 1101..1126
FT /note- "THE TNT repeat 3"
FT 1158..1183
FT /note- "THE TNT repeat 4"
FT 1286..1311
FT /note- "THE TNT repeat 5"
FT 1314..1339
FT /note- "THE TNT repeat 6"
FT 1349..1374
FT /note- "THE TNT repeat 7"
FT 1414..1439
FT /note- "THE TNT repeat 8"
FT 1774..1781
FT /note- "peptide 293 2nd sequence"
FT 1808..1819
FT /note- "peptide 115"
FT 1819..1840
FT /note- "peptide 261 1st sequence"
FT 1853..1863
FT /note- "peptide 240"
FT 1901..1919
FT /note- "peptide R32"
FT 1919..1930
FT /note- "peptide 261 2nd sequence"
XX
PN MO9413315-A.
XX
PD 23-JUN-1994.
XX
PF 03-DEC-1993; 93MO-US11721.
XX
PR 04-DEC-1992; 92US-0989842.
PR 12-APR-1993; 93US-0046585.
XX
PA (COLD-) COLD SPRING HARBOR LAB.
PA (TULA-) TULARIK INC.
XX
PI Herr W, Lamarco K, Wilson A;
XX WPI; 1994-234207/28.
DR N-PSDB; AAO69229.
XX
PT New Host Cell Factor polypeptide(s) and nucleic acid - are used
PT to develop agents for diagnosis or treatment of disease
PT associated with expression of a HCF-modulated gene e.g. viral

PT infections
XX
PS Disclosure; Page 39; 71pp: English.
XX
CC HCF is required for the transcription of a number of
CC viral genes, such as the immediate early herpes simplex virus-1
CC genes. Epitopes of the encoded protein can be used in
CC defining functional domains of HCF, identifying compounds that
CC associate with HCF or designing compounds capable of modifying HCF
CC transcription. Such agents can be used to treat viral infections.
XX
SQ Sequence 2035 AA;
Query Match 5.6%; Score 237; DB 15; Length 2035;
Best Local Similarity 22.6%; Pred. No. 1.1e-06;
Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;
OY 10 VSAPKVSQPRLPAPQIVAAKAPRTTITIOFPANFOLPPGVILKINSGLMLVSPQOTV 69
DB 537 IGSSPQMSGMALAAAAATOKIPSSA---PTVLSPAGCTIVKT---NAVTPGTTT 588
OY 70 TRAEITSNITSRPAPVANDPQVTKICTVNSSOLIKVAVPVKLAIGTTV----- 122
DB 589 LPA--TVKVASSPVAVSNPAT-----RMKRTAA-----AOVTSVSSATNTS 628
OY 123 ---VITVPKPSVQSVAVPTSVVTVPGKPLNTVTLK--PSSL--GASSTPSNEPML---- 173
DB 629 TRPITTVHKSGTV--IVAAQAAQVTTVVGAVFTITILVSPISVPGSALISMKGVMVY 687
OY 174 --KAENSAVQINLSPTMLENVKCKKNFL--AMLIKLACSGSQSP-----EMGQNVK 221
DB 688 QTKPVQTSAVVQGAQSTGVTQIITQKCPPLPAGTIILKVLTSAGKCPPTIITTTQASGAGK 747
OY 222 KLVEQLDLAKIEAEFTKRLVLELKSPOPHVPLPKSVVALRDLNNSQFIQOCVQ 281
DB 748 PTILGI-----SSVSPSTT---KCGTTTITITPMSALITTAAGATG 785
OY 282 TSSDMVIACTTTT---SPVVT-----TVSSSOSKESIIVSGA-----TA 321
DB 786 VTSSPGIKSPITITITKMTSGTGAPAKIITAVPKIATGHGQGVTVYVLKAGAPQPGTI 845
OY 322 PRT-----VSQTLNPLAGPVGAK--AGVVTLSHSGPTAAGCTTAGTLLOTS 368
DB 846 LRTVPMGCVRLVTPVTVSAVAVKPAVTVTLVKGTTGVTTLGTVGVST--SLAGAGCHSTS 903
OY 369 KPLVTSVA--NTVTVSLOPEKPVVSGTAVTLSTLPAVTFEGTSGAALCLPSVKPVVSCW 426
DB 904 ASLATPITTLGTLATLSSQ---VINPRAITVSAAQTLTLAAGGLTPTTMOV----- 954
OY 427 DHICKPVIGTPOIKL-AOPGVLSQPA-GIPTGSSSKQLFSLFHVVOQPSGNGEKOVTT 484
DB 955 -----SQPLQVTLINAPSGVENQPVHDLVY-----ILASP-----TT 987
OY 485 ISHSSTLTIOKCGKQKTMVNTIIPTSOPPPASI-----LKQITLPGNKILISLQASPTQ 537
DB 988 EOPFATVITADSGQGDVOPGVTLVCSNPCEHETHTTNTATTVVAN--LGHPQPTQ 1045
OY 538 KNRIKENTSCFDEDDINDVTSAGVNLNENNAOILATNSLVOTLQSCDEP 592
DB 1046 VQFV-----CDROEAAASLVITVG---QDN-----GSVVAVCSNPP 1079

Search completed: February 16, 2003, 21:54:38
Job time : 83.3139 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 16.7846 Seconds

(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GILVTVAPVAPVAPKVSQGP.....KZHOHERALFTIRLTLLTVY 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 2554876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCY_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/PCYUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1157.5	27.1	9 US-09-893-519A-14	Sequence 14, Appl
2	220	5.2	10 US-09-801-368-108	Sequence 108, App
3	219	5.1	9 US-10-025-380-1068	Sequence 1068, Ap
4	219	5.1	10 US-09-922-217-1068	Sequence 1068, Ap
5	219	5.1	10 US-09-833-263-1068	Sequence 1068, Ap
6	203.5	4.8	10 US-09-801-368-114	Sequence 114, App
7	184.5	4.3	10 US-09-801-368-104	Sequence 104, App
8	182.5	4.3	10 US-09-862-027-40	Sequence 40, App
9	182	4.3	10 US-09-801-368-110	Sequence 110, App
10	182	4.3	10 US-09-864-761-34248	Sequence 34248, A
11	172	4.0	10 US-09-801-368-106	Sequence 106, App
12	167	3.9	9 US-10-161-510-10	Sequence 10, Appl
13	159.5	3.7	10 US-09-905-129-21	Sequence 21, Appl
14	159.5	3.7	10 US-09-991-630-21	Sequence 21, Appl
15	158.5	3.7	10 US-09-905-129-2	Sequence 2, Appl
16	158.5	3.7	10 US-09-905-129-10	Sequence 10, Appl
17	158.5	3.7	10 US-09-905-129-13	Sequence 13, Appl
18	158.5	3.7	10 US-09-991-630-12	Sequence 12, Appl
19	158.5	3.7	10 US-09-991-630-10	Sequence 10, Appl

20	158.5	3.7	2597	10 US-09-991-630-13	Sequence 13, Appl
21	157.5	3.7	941	12 US-10-124-557-14	Sequence 14, Appl
22	157.5	3.7	1022	12 US-10-124-557-84	Sequence 84, Appl
23	157.5	3.7	1038	12 US-10-124-557-74	Sequence 74, Appl
24	157.5	3.7	1049	12 US-10-124-557-58	Sequence 58, Appl
25	157.5	3.7	1140	12 US-10-124-557-104	Sequence 104, App
26	157.5	3.7	1270	12 US-10-124-557-44	Sequence 44, Appl
27	157.5	3.7	1311	12 US-10-124-557-42	Sequence 42, Appl
28	157.5	3.7	1313	12 US-10-124-557-142	Sequence 142, App
29	157.5	3.7	1314	12 US-10-124-557-50	Sequence 50, Appl
30	157.5	3.7	1320	12 US-10-124-557-46	Sequence 46, Appl
31	157.5	3.7	1320	12 US-10-124-557-60	Sequence 60, Appl
32	157.5	3.7	1354	12 US-10-124-557-48	Sequence 48, Appl
33	157.5	3.7	1361	12 US-10-124-557-40	Sequence 40, Appl
34	157.5	3.7	1363	12 US-10-124-557-52	Sequence 52, Appl
35	157.5	3.7	1404	12 US-10-124-557-2	Sequence 2, Appl
36	157.5	3.7	1404	12 US-10-124-557-62	Sequence 62, Appl
37	157	3.7	594	10 US-09-801-368-112	Sequence 112, App
38	157	3.7	688	10 US-09-864-761-36047	Sequence 36047, A
39	156	3.7	2586	10 US-09-905-129-11	Sequence 11, Appl
40	156	3.7	2586	10 US-09-905-129-14	Sequence 14, Appl
41	156	3.7	2586	10 US-09-905-129-11	Sequence 11, Appl
42	156	3.7	2586	10 US-09-991-630-14	Sequence 14, Appl
43	156	3.7	2587	10 US-09-905-129-16	Sequence 16, Appl
44	156	3.7	2587	10 US-09-991-630-16	Sequence 16, Appl
45	156	3.7	2589	10 US-09-991-630-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-09-893-519A-14
Sequence 14, Application US/09893519A
Publication NO. US20030027243A1
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BUTRMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAQ, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OR INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G348-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 1023
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
DATABASE ENTRY DATE: 1997-06-25
RELEVANT RESIDUES: (1)..(1023)

Db 947 SKSVIIPSS--STGSGSESTGSSASSSSSSSSSPKSTYSSSLPPTSATSGEI 1004
Qy 397 TLSIPAVFEGTGAICLPVKKPVSEPCMDHICKPVIGTPV---QIKLQPGVLSQP 452
Db 1005 TSSLPVTTTTSQETTLV-----TVTSCESHVCTESISSAIVATYVSGATTEYTTW 1059
Qy 453 AGIPTGSSSKQLFSLFHVQOPSGCNEKQVTTISHSSTLTQKC-----GQ 498
Db 1060 CPTSTTEITKQ---TTEETTKQTKGTTEQTEETTKQTVVTTISCSBDCSKTASPAIVST 1116
Qy 499 KTMVNTI-----IPTSQFPASILKQITLPGNKILSLQSPYCKNKIKENVTCF 549
Db 1117 STATINGTTEYTWCPSTSTESKQOQTLTVTSCSGVCSGETTSPA---IVSTATA-- 1170
Qy 550 RDEDINDVTSAGVNLNEENACILATNSELVGLIOSCKDEPFLIGALQ----- 600
Db 1171 ---TVNDVTVVSTWRQ-----TTMEQSVSKMNSATSETTNTGAETTTSGAAE 1220
Qy 601 -----KRILDGKHHDTTELNSDAVNLISQATQERL 631
Db 1221 TKTVVTSISRFNHAETQATADVYIGHSSSVSVSEGTGNKLSLSSGLSTMSQ--QPRS 1278
Qy 632 RGLLEKL-TATAQHRMTYKASENTIL 657
Db 1279 TPASSWGSSTASLEISTYAGSANSLL 1305

RESULT 7

US-09-801-368-104
Sequence 104, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amlr
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 104
LENGTH: 1537
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-104

Query Match 4.3%; Score 184.5; DB 10; Length 1537;
Best Local Similarity 19.3%; Pred. No. 0.0011;
Matches 134; Conservative 104; Mismatches 257; Indels 201; Gaps 27;

Qy 50 TVLI---KSNGLPLMYSPOOTVTRAETTSITSRPVPANP--QTVKICVPPNSSQLI 104
Db 938 TVIIRPTSGLISTTEPTGTSTSTEVTTITGTCQPTDETVIIRPTSE--- 993
Qy 105 KKVAVTPVKLAQIGTVTVTPKPSVOSVAVPTSVTVPGKRLN-TVTTIK-PSSLG 162
Db 994 -----GLISTTEPTGTSTSTEVTTITGTCQPTDETVIIRPTSEG 1039

Qy 163 ASSTPSNEP-----NLKENSAAVOINLSPMLENVKCKNFLMLIKLACSGSQSPEN 216
Db 1040 LVTT-TTEPTMGTFTSTSTEMSTVCTGNLPLD-ETVIVVAT-PTTALSSSLSSSQOI 1096
Qy 217 GONVAKLYEQLDAKIEAEFTFKRLYELKSSPOPHLVPELKSVALROLPLNSQFTQ 276
Db 1097 TSSI-----TSSRPIIRPFYPS-----NGTSVI- 1119
Qy 277 QCVQOTSSDMVIANCTTV-TTSPVVTTVSS-----SOSERSIIV-----SGA 319
Db 1120 -----SSVSISSVTSLSFTSSPVISSVTSSTTSTISFESSKSVIPTSSTSGS 1173
Qy 320 TAPRTVSQTLNPLAGPVAKAGVYTLHSGVTAATGTTAGTGLQTSKPLVSVANTV 379
Db 1174 SESESTSS-----AGSVSSSSSFISSESKSPYSS-----SLPLVTSATTSQ 1215
Qy 380 TVVSLQPKPVVSGTAVTSLSPAVTFEGTSGAICLPVKKPVSEPCMDHIC---KPVIG 435
Db 1216 ETAS-----SLPPTATTKTSQETTLV-----TVTSCESHVCTESISPAIV 1255
Qy 436 TPVQIKLQPGVLSQPAGIPTGSSSKQLFSLFHVQOPSGCNEKQVTTISHSSTLTQK 495
Db 1256 STATVTV-----SGVTEYTWCPSTTEETTKQTKGTTEQTEETTKQTVVTTISS 1305
Qy 496 C-----GQKTMVNTI-----IPTSQFPASILKQITLPGNKILSLQ 532
Db 1306 CESDVCSTASPAIVSTGATINGTTEYTWCPSTSTESKQOQTLTVTSCSGVCSGET 1365
Qy 533 ASP-----TQKNRIKENVT-----SCFRDEDINDVTSAGVNLNEENACILATNSELV 581
Db 1366 ASPAIVSTATATVNDVTVVPTWRPQTANESVSSKMSATGETTNTLAAETTTVAA 1425
Qy 582 GTLIQSCKDEPFLIGALQKR-----ILDIGKHHDTTELNSDAVN 621
Db 1426 ETTTNGALETKTVYVTSLSNSNHAETQATADVYIGHSSSVSVSEGTGNKLSLSSGLS 1485
Qy 622 LISQATQERLGLLEKL-TATAQHRMTYKASENTIL 657
Db 1486 TMSQOPRSTPASSWVGYSST-ASLEISTYAGSANSLL 1520

RESULT 8

US-09-862-027-40
Sequence 40, Application US/09862027
Patent No. US20020142428A1
GENERAL INFORMATION:
APPLICANT: Hodge, Martin R.
TITLE OF INVENTION: No. US20020142428A1el Kinases and Uses Thereof
FILE REFERENCE: 35800/234862
CURRENT APPLICATION NUMBER: US/09/862,027
CURRENT FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 09/345,473
PRIOR FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 82
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 40
LENGTH: 1601
TYPE: PRT
ORGANISM: C. elegans
US-09-862-027-40

Query Match 4.3%; Score 182.5; DB 10; Length 1601;
Best Local Similarity 19.4%; Pred. No. 0.0016;
Matches 136; Conservative 115; Mismatches 293; Indels 157; Gaps 29;

Qy 5 TKVAVSAPRVVSSGPRLLPAPQIVAVKAPNTTITFQPANLQLPQTVLIRKNSGPIMLVS 64
Db 514 TPIPTPATPH-SSAQOQPIPPPLSTQ---TSAEIQGSAQOP-----SVPVTMIA 559
Qy 65 ---POOTVTRAETTSITSRPVPANPQIVKICVPPNSSQLIKKVAVTPVKLAQIGTV 122
Db 560 NIPAMSPSAPQPVLS--PTSAAVPVTMTIHVKPSEIPQVNAVTT---AAVAVANN 613

```

OY      123 VTTVKKPSSVSAVPE-----SVTVTPRGKPLNVTL-----KPSLGSAGS-----165
|      |      |      |      |      |      |      |      |      |      |      |
Db      614 VPPSPAPRKTEIDIQPTLQONTVPRTISTDASGLVINTPASIASBPAPSPATDVASTTAP 673
|      |      |      |      |      |      |      |      |      |      |      |
OY      166 -TPSPNEPLIKANSAVAOINLSPTMLENVKCKKNEFLMLIKLACSGSQ-----SPENGON 219
|      |      |      |      |      |      |      |      |      |      |      |
Db      674 VTPADTPPTTDDGAAASSTTENKEEKRKGNKRKVMEI-LGDDESRNFALVSCRIDTS 732
|      |      |      |      |      |      |      |      |      |      |      |
OY      220 VKKLEQL-----LDAKIEAEFTFKLIVELKSSPOHLYPFL-----KKSV-VALR 265
|      |      |      |      |      |      |      |      |      |      |      |
Db      733 HKSVTFQAPGTDRKCTTATKLAEDECLKHVHIVEQLQVEIQLINSDGKKGVTKLA 792
|      |      |      |      |      |      |      |      |      |      |      |
OY      266 QLL-PSNQS-FIQOCVQOOTSMDVIATCTTYTSPVVT--TVSSSQSEKSIYSGAT 320
|      |      |      |      |      |      |      |      |      |      |      |
Db      793 TVLDONSTPEPTITAVMKDSS-----AATASMTKRIEIEKTPPRDASQEBNNVQV---T 846
|      |      |      |      |      |      |      |      |      |      |      |
OY      321 APRTVSVOTLNLPLAGPVGAKAGVTLHVSVPPTAAITGGTTACTG-----LIQTSKPL 371
|      |      |      |      |      |      |      |      |      |      |      |
Db      847 NVRKVSQESNMAESVQISIRPRGIIYM--SEPTNOTDASAPPTGAAPKSRFOYTKSADPI 903
|      |      |      |      |      |      |      |      |      |      |      |
OY      372 VTSVANITYTYSLOEKRVVSGTAVTLSLPAVTCGTSGAALICPSKRPVVSFCWDHICK 431
|      |      |      |      |      |      |      |      |      |      |      |
Db      904 ATPISSISSTATV--IPVATPTNI-----TSEPIVOPITAOVYT-----943
|      |      |      |      |      |      |      |      |      |      |      |
OY      432 PVIGTPVOIKLAQPGV---LSOPAGIPTGSSSKOLFSLFHVVOQPSGAGNEKOVTTISHS 488
|      |      |      |      |      |      |      |      |      |      |      |
Db      944 -----HLATPSRPSHSLSSNMS-PSATTHSNMSI-----QSTTVSPG 980
|      |      |      |      |      |      |      |      |      |      |      |
OY      489 STLTIQCKGQKTPMVTNIITPSQPPASILKQITLP-----GNKILSLQASPTQKNR 540
|      |      |      |      |      |      |      |      |      |      |      |
Db      981 RRTVQVPVQAQESGSISSITPHPEPTPAITSCPPVPVSVPPVNSGTLNLEVAQK---1037
|      |      |      |      |      |      |      |      |      |      |      |
OY      541 IKENTSCFREDIDINDVTSAGVNLNENMCILATNSLVTGLQSCKDEPFLFALQ 600
|      |      |      |      |      |      |      |      |      |      |      |
Db      1038 -----TPSATNONVDTQSHSSSTASTATLSETPATVHTPISVPAPVQ 1080
|      |      |      |      |      |      |      |      |      |      |      |
OY      601 KRIDIGKKHDTITELNSDAVNLISATOERIGLLEKLTAI 641
|      |      |      |      |      |      |      |      |      |      |      |
Db      1081 EPLVIDHHSVDVLTQLDSELKRVSGVSHSPSTYVESLISM 1121
|      |      |      |      |      |      |      |      |      |      |      |

RESULT 9
US-09-801-368-110
; Sequence 110, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amit
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 110
; LENGTH: 1075
; TYPE: prt

```

```

: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 34248
: LENGTH: 2665
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL034555.2
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
: OTHER INFORMATION: EST_HUMAN HIT: AUL17052.1, EVALUOE 0.00e+00
: OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUOE 3.00e-10
US-09-864-761-34248
```

```

Query Match 4.3%, Score 182, DB 10, Length 2665,
Best Local Similarity 22.2%, Pred. No. 0.0035,
Matches 133, Conservative 76, Mismatches 230, Indels 160, Gaps 29;
```

```

Qy 9 PVSAPKTVSSGRLPAQIVAKPANTTTIOFPANLQLPFGVILKSNGLMYLS--PQ 66
Db 1841 PVTTP-----SDPSITPIPLPVTA-----AKLSPPVASGGIPHOSSPPKVTETWTR 1887
Qy 67 QTVTAETTSNITSRPAPVAPNQTIVKICTVPSNSQLIKKVAVT--VKLAQIGTTVVT 124
Db 1888 QEPPRAQSTPS-----PALPDPDKASDVDRSSSTL-RKILMDPKYVSATVSTSTVTT 1939
Qy 125 TVPKRSSVQSAVAVPTSVVTTPGKPLNVTYTLKPSLSGASTPSNEPNIKASNAVAQIN 184
```

```

Db 1940 AIAEPVSAAPCLHEAPPPPVDSKKPLEEKTA--PPVTNNSEIQASEVLVAADKEKAVAPV- 1996
Qy 185 ISPTMLENVKCKKCNFLAMLIKACSGSOSPENGQVKKVLEOLDPAKIEAEFTKRLYE 244
Db 1997 IAPKITSYISR-----MPVSIDLENSQKITLAKAPQILITLVA----- 2036
Qy 245 LKSSDPHLVPLPKKSVYALROLPLPNSOSFIOOCVQOOTSDDMVIACTCTTYYTTS--VVT 302
Db 2037 LGLVNVSLVP-----VNALKGPVKGVSYYTLKSLVSTFAGPVNVLKGPVNVLTGVVNI 2091
Qy 303 TTVSSSSEKSLIIVSGAIAKPTVSVQTLNPLAGPVGAKAGVYTLASVGFPAATGCTTACT 362
Db 2092 TPNVAT-----VGTVNAAPGTV-----NAAASAVNAAASAVTAVGAVTAASGGVTAAT 2140
Qy 363 GLLQTSKPLVTSVANTVYTVSLQPEKPVVSGTAVTSLPVTAFEGTSGAICLPPSVKPVV 422
Db 2141 GTV----- 2155
Qy 423 SPCWDHICKPVIGTPVQIKLAQPG--PVL-SOPAGIPGSSSKQLPSLPHVVOQPSGNE 479
Db 2156 -----CKQRASANENSRR-HPGSMPIYIDRRPADAGSGAGLR-----VNTSEG--- 2196
Qy 480 KQYTTISHSSTLTQKCGQKTM-PVNTTIPTPSQFPASTLKQITTLPGKILSLQASPTQK 538
Db 2197 -----GQKTEGPORISAKISQIPASAMD-----IFEQOS-VSK 2236
Qy 539 NRK-ENYTSCEPDEDIDVTSMAG-VNLNENACILATNSELVTLIOSCK-DEPFL 594
Db 2237 SQVKPDSTVA-----SQPPSKGPQAPGAYANVATHSLVLTQATYNNASPVYSSVKADRPFL 2292
```

```

RESULT 11
US-09-801-368-106
: Sequence 106, Application US/09801368
: Patent No. US20020128250A1
: GENERAL INFORMATION:
: APPLICANT: Busby, Robert
: APPLICANT: Call, Brian
: APPLICANT: Hecht, Peter
: APPLICANT: Holtzman, Doug
: APPLICANT: Madden, Kevin
: APPLICANT: Maxon, Mary
: APPLICANT: Milne, Todd
: APPLICANT: No. US20020128250A1man, Thea
: APPLICANT: Royer, John
: APPLICANT: Salama, Sofie
: APPLICANT: Sherman, Amir
: APPLICANT: Silva, Jeff
: APPLICANT: Summers, Eric
: TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
: FILE REFERENCE: 109272.147
: CURRENT FILING DATE: 2001-03-07
: PRIOR FILING DATE: 2001-03-07
: PRIOR APPLICATION NUMBER: US 09/487,558
: PRIOR FILING DATE: 2000-01-19
: PRIOR APPLICATION NUMBER: US 60/160,587
: PRIOR FILING DATE: 1999-10-20
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 106
: LENGTH: 1169
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106
```

```

Query Match 4.0%, Score 172, DB 10, Length 1169,
Best Local Similarity 21.4%, Pred. No. 0.0047,
Matches 112, Conservative 85, Mismatches 191, Indels 136, Gaps 24;
```

```

Qy 4 YTKVAPVSAAPPVSSGRLPAQIVAKPANTTTIOF-PANLQLPFGVILKSNGLPLML 62
Db 689 YTSSEVSTPIITSESSASVYTL-----PSITSEFKPSTWK-----TKVYSSSSPTNL 740
```

QY 63 VSPQQTATRAETSNITSRAY-----PANPOTVACTVPSNSQOLIKKAVNTPV 112
DB 741 ITSYDTTSSKDSVSTSVSLSSISLPSYSASSSBOIHRSSISVNGALTSFSSSTKV 800
QY 113 KK-----LAQIGTVVTVTPK-----SSVQSVAVPTSVTVTPGPK 149
DB 801 SSSSSSSHTSPITSSSESIKSGVIEESTSTSSSFHETSTASTSVQJSSQVTTSSP 860
QY 150 LNTVTLTKPSSLC-----ASSTPSNEPMLKAENSAVAOINLSPTMLENVKCKNLAFLIK 205
DB 861 ISTVA---PSTGLNSQFESTSNSSKETMSSENSASV-----893
QY 206 LAGSGSPMGONKKVLEQLDIAKEASE-----FMR---KLYVELKSSPOHLYPFLK 256
DB 894 MPSSSATSPTG-----KVTSDETSSEFSRDRRTTVYRMTSEPTSN---EQ 936
QY 259 KSVVALROLPNSSQFIQOCVQOQSSDMVATCTTV-----TT---SPVTTTVSS 307
DB 937 TLLITVSSCSNSGS-----MTVSSAVVSTATTINGITTEYTCPLSATLTVSK 989
QY 308 SQS-EKSIIVSGATPPTVVOQLNP--LAGPVGAKGVYTLHSGVPTAATGCTAGTGL 364
DB .990 LSEBEKTLITVYSCSEGVSETPASPAIVSTAVTAVDVTVYSTMSPQATNKLAVSSD- 1048
QY 365 LOTSPLVTSVANTVTVYSLQPEKPVYSGTAVTSLSLPAVTFEGSGAICLPYKPVVSF 424
DB 1049 IENSASASVSEAEATKTSIRNNNFVP-TSGTTSIETHP-TTTSNSSENSDNVS----- 1101
QY 425 CMHDICKPVIGTPVOIKLAOPGVLSQAPGIP-----GSSSKQI 464
DB 1102 ASEAVSSKSVTNPVLSVSQ-----QPRGTASSMIGSSTASL 1139

RESULT 12
US-10-161-510-10
; Sequence 10, Application US/10161510
; Publication No. US20020192695A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: PIBS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-074C
; CURRENT APPLICATION NUMBER: US/10/161,510
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1056
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-161-510-10

Query Match 3.9%; Score 167; DB 9; Length 1056;
Best Local Similarity 21.4%; Pred. No. 0.0086;
Matches 134; Conservative 61; Mismatches 238; Indels 192; Gaps 24;
QY 7 VAVVSAAPKVSQGRRLPAPDIVAKAANTTTIOFPANLQLPFGTVLIXNSGP-LMLVSP 65
DB 39 VAOGTGAPSKVDSSFOLEPAKKNAAL-GSEPRRL---ALAPVGPRAAKMSASSEGRLALASP 94
QY 66 QQTVTRAETSNITSRP--AVPANPQTVKICTVPSNSQOLIKKAAVVPVKKLAQIGTTV 123
DB 95 -----RPIALPCTPEGOKATATHRSS-----LAPISVOL--VMSASA 132
QY 124 TTVPKSSVQSVAVPTSVTVTPGK--PLNTVTLTKPSSIGASSTPSNEPMLAENSAV 181
DB 133 GPKPPTATGSLAPTSIGLIMPASAGPRSPVTLGPNLAPTSRDQOEP-----PA 184

QY 182 QINLSPTMLENVKCKNLAFLIKLACSGSGSPMGONKKVLEQLDIAKEAETTRKL 241
DB 185 SVGKPPYL-----AASGLIALASEROP-----208
QY 242 YVELKSSPOHLYPFLKSSVALNQLLPNSQFIQOCVQOQSSDMVATCTTVTSPV 301
DB 209 --ELPSTPSP--VP-----SPVL 222
QY 302 TTVYSSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKGVYTLHSGVPTAATGCTAG 361
DB 223 SPQEQALAPASFTASGAASVQGTARKRDAAPARPLPASBG-----HLQPAQSGPTGS 277
QY 362 TGLLOTGKRP--LYTSVANTVTVYSLQPEKPVYSGTAVTSL-----P 401
DB 278 PRCIQTSPDRLPSFSFRAPREALHSSPEDDVLPRPQTLPLDVGQGPSEGTSPGLLSP 337
QY 402 AVTFEGTSGAICLPYKPVVSFCMDHICKPVIGTPVOIKLAOPGVLSOP-AGIPGSS 460
DB 338 TFRGASGQGVTPPLPKPRPS-----PSRPSHSPPKSPCVPAPDMALPRLGTOSTGP 392
QY 461 SKOLFSLPHVVOQSPSGNEKQVTTISHSSTLTQKCGOKTAPVNTTIPTSOFPASILKQ 520
DB 393 GRCLSPMLAQOEAFA-----PVTSSSTSTLSSSPMSAQ-----PTMKSDPGRITV 439
QY 521 IT-----LEGKTLISQASPTQKNRIKENVTSQFREDINDVTSAGVNLAEENACIL 574
DB 440 VTNWNVGTAMPDDVTSL-----LHLGGDDSDGADMAIGLQEVNSML- 482
QY 575 ATNSELVGTILQSCDEPLFICAL 599
DB 483 --MKRLDALFTQWSE--LEMDAL 503

RESULT 13
US-09-905-129-21
; Sequence 21, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Elnat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS TH
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 2828
; TYPE: PRN
; ORGANISM: homo sapiens
US-09-905-129-21

Query Match 3.7%; Score 159.5; DB 10; Length 2828;
Best Local Similarity 20.4%; Pred. No. 0.1; Mismatches 241; Indels 245; Gaps 30;
Matches 144; Conservative 77;
QY 50 TVLIKSNCGPLMLVSPQQTATRAETSNITSRPAVAPNPQTVKICVVPN-----SSQL 103
DB 1053 TLLIKKMKKMSQTLQGNLMLEGDPTHSSE-----SEGOESSTILPDSLTGIMSMSP 1108
QY 104 IKKVAVTPVKKLAQIGTTVTVPKSSVQSVAVPTSVTVTPGK-----148
DB 1109 VKKPAETTVGTLDKDTTVTTTPR-----QKVA--PSTSMSTHRSRRPNRRRLRPKKR 1163
QY 149 ---PLNTVTLTKPSSIGASSTPSNEPMLKAENSAVAOINLSPT-----MLENVK 194

```
Db 1164 HRHKQPTPTFPABSET-FSTQPTQAPDIKI--SSQVSSSLVPTAMVDNTVNTPKOLEMEK 1220
Qy 195 KCR-----NFLAMLIKLASGSGSOPREMGONVK-----LVEQ 226
Db 1221 NAEPTSGKTRRRKHGRKPNKHRTPTSVSSRASGSKSPSPENKHRIIVTPSSFTILLPR 1280
Qy 227 LLDKATIAE-----EFTRLVYELKSSPO-PHLVFLKSSVVALROLLPNSQSF10QC 278
Db 1281 TVSLKTEGYPYDSLDMYMTTKRIY--SSYPKVOEFTLPVTKPPTSDGKEI---KDVATN 1333
Qy 279 VQQTSSSMVIA--TCCTVTVTSPVVTYSSSGSEKSIIVSGATAPRTVSQTLNPLAGP 336
Db 1334 VDKHRSIDLVTGESITNAIPISRLSVSTMGFEKSESS-----P 1371
Qy 337 VGAKAGVVLHVSVPATAGTGTAGTGLQTSKPLVTSVANVTYVLSLOPEKPYVSGTAV 396
Db 1372 VGFP-----GTPWNSRRTAQPRLOTIDPVTTSGEN-LTDPPLAKELEVDFTSE 1421
Qy 397 TLSLPAVT---FGETSGAAILCPSPVKPVY-----SFCWDHI-----CKPVIG 435
Db 1422 FLSSLVSTFPHQEAGSSFTLSSIKVEVASSQAETTTLLDQDHEFTVAILLSETRPNH 1481
Qy 436 TPVQIKLAOG-----PYLSOPAGIPFGSSSKQLFSLFHV-----470
Db 1482 TPTAARKKEPSSPSTILMSLGOTTTKTPALPSPRISQASRDSKENVFLNYGNPETEA 1541
Qy 471 -----VOOPSGNE-----KQVTTISHSSTLTIOKGOKTM-----501
Db 1542 TPVNNECTQHMSPNEIESTSSDRAFNLSKLEKQVFGSRLPFGPDSORODGAVHA 1601
Qy 502 -----PVNTIIP-----TSQPP-----PASIL---KQ 520
Db 1602 SHQLTRVAPKPIPLTAVRLPEMSTQASRYFVTSQSPRHMTNKPEITTYPSGALPENKQ 1661
Qy 521 ITLPGKILISQASPTQKNRIKENVTSQFRDE--DDINDVTSNAGVN 565
Db 1662 FTTPR---LSSTTIPPLPHMSKPSIPSKFTDRIDQNGSKYVGN 1705
```

```
RESULT 14
US-09-991-630-21
; Sequence 21, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
; APPLICANT: Elnat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THEREOF
; FILE REFERENCE: 540579-2007.3
; CURRENT APPLICATION NUMBER: US/09/991,630
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-991-630-21
```

```
Query Match 3.78; Score 159.5; DB 10; Length 2828;
Best Local Similarity 20.4%; Pred. No. 0.1;
Matches 144; Conservative 77; Mismatches 241; Indels 245; Gaps 30;
```

```
Qy 50 TVLIKNSGCLMLVSPQQTVAETTSNISRPAVPANPQTVKICVYPN-----SSQL 103
Db 1053 TLIIKGMKMKMSQTLQGGNMLEGDPHRSRSE-----SEGESKSTYLPDSTLGMSSMSP 1108
```

```
Qy 104 IKKAVTPVKKLAQIGTIVTTPKPSVSVAVPTSVNTVTPGK-----148
Db 1109 VKKPAETFTVGLDKDITVTTPR-----QKVA-PSTMSHPRRRPNGRRLRPNKR 1163
Qy 149 ---PLNTVTLKPSLASSTPSPNEPMLKAENGAAYINLSPT-----MLENVK 194
Db 1164 HRHKQPTPTFPABSET-FSTQPTQAPDIKI--SSQVSSSLVPTAMVDNTVNTPKOLEMEK 1220
Qy 195 KCR-----NFLAMLIKLASGSGSOPREMGONVK-----LVEQ 226
Db 1221 NAEPTSGKTRRRKHGRKPNKHRTPTSVSSRASGSKSPSPENKHRIIVTPSSFTILLPR 1280
Qy 227 LLDKATIAE-----EFTRLVYELKSSPO-PHLVFLKSSVVALROLLPNSQSF10QC 278
Db 1281 TVSLKTEGYPYDSLDMYMTTKRIY--SSYPKVOEFTLPVTKPPTSDGKEI---KDVATN 1333
Qy 279 VQQTSSSMVIA--TCCTVTVTSPVVTYSSSGSEKSIIVSGATAPRTVSQTLNPLAGP 336
Db 1334 VDKHRSIDLVTGESITNAIPISRLSVSTMGFEKSESS-----P 1371
Qy 337 VGAKAGVVLHVSVPATAGTGTAGTGLQTSKPLVTSVANVTYVLSLOPEKPYVSGTAV 396
Db 1372 VGFP-----GTPWNSRRTAQPRLOTIDPVTTSGEN-LTDPPLAKELEVDFTSE 1421
Qy 397 TLSLPAVT---FGETSGAAILCPSPVKPVY-----SFCWDHI-----CKPVIG 435
Db 1422 FLSSLVSTFPHQEAGSSFTLSSIKVEVASSQAETTTLLDQDHEFTVAILLSETRPNH 1481
Qy 436 TPVQIKLAOG-----PYLSOPAGIPFGSSSKQLFSLFHV-----470
Db 1482 TPTAARKKEPSSPSTILMSLGOTTTKTPALPSPRISQASRDSKENVFLNYGNPETEA 1541
Qy 471 -----VOOPSGNE-----KQVTTISHSSTLTIOKGOKTM-----501
Db 1542 TPVNNECTQHMSPNEIESTSSDRAFNLSKLEKQVFGSRLPFGPDSORODGAVHA 1601
Qy 502 -----PVNTIIP-----TSQPP-----PASIL---KQ 520
Db 1602 SHQLTRVAPKPIPLTAVRLPEMSTQASRYFVTSQSPRHMTNKPEITTYPSGALPENKQ 1661
Qy 521 ITLPGKILISQASPTQKNRIKENVTSQFRDE--DDINDVTSNAGVN 565
Db 1662 FTTPR---LSSTTIPPLPHMSKPSIPSKFTDRIDQNGSKYVGN 1705
```

```
RESULT 15
US-09-905-129-2
; Sequence 2, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Elnat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THEREOF
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2597
; TYPE: PRT
; ORGANISM: rattus species
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(2597)
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 28.8257 Seconds
(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264
Sequence: 1 GLVTRVAPVAPVAPPKVSQGP.....KZHOHERALFTIRLTLLTY 852

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	775	18.2	921	2 A48184	transcription init
2	775	18.2	921	2 A45183	TRP-associated fac
3	344	8.1	549	2 C87719	protein R19.6 [tm
4	244.5	5.7	2187	2 T30826	nascent polypeptid
5	237	5.6	2035	2 A40718	host cell factor C
6	234.5	5.5	3020	2 A43932	mucin 2 precursor,
7	233.5	5.5	528	2 I47141	gastric mucin (clo
8	230.5	5.4	1275	2 T33369	hypothetical prote
9	220	5.2	1367	2 A48478	glucan 1,4-alpha-g
10	213.5	5.0	5376	2 T42215	zonadhesin - mouse
11	212.5	5.0	2232	2 T34434	hypothetical prote
12	210.5	4.9	1777	2 T34369	hypothetical prote
13	210	4.9	873	2 A47283	calphostin - fruit
14	206.5	4.8	865	2 A47282	calcium-binding pr
15	206	4.8	1032	2 T34433	hypothetical prote
16	205.5	4.8	1161	2 S57180	probable membrane
17	204.5	4.8	1324	2 S52863	DNA-binding protei
18	203.5	4.8	873	2 F96615	probable MyD-famil
19	203	4.8	697	2 A86402	protein T22C5.17 [
20	201	4.7	725	2 A41258	a-actinin core
21	200.5	4.7	1367	2 S51959	hypothetical prote
22	200	4.7	3507	2 T34513	hypothetical prote
23	198	4.6	867	2 T3463	membrane glycoprot
24	197.5	4.6	881	2 S56032	probable membrane
25	197.5	4.6	1609	2 S25345	mucin 3 - human (f
26	197	4.6	648	2 PC4395	hypothetical prote
27	193.5	4.5	670	2 T36791	hypothetical prote
28	193	4.5	825	2 T29634	hypothetical prote
29	192	4.5	1229	2 T25697	hypothetical prote

30	192	4.5	3759	2 A35085	tilthorax protein
31	191	4.5	797	1 VGBEX1	glycoprotein X pre
32	190.5	4.5	1260	2 S60896	agglutinin-like pr
33	189.5	4.4	851	2 T22696	hypothetical prote
34	189.5	4.4	866	2 T45462	membrane glycoprot
35	189.5	4.4	1630	2 A53577	ascleris sialoglyco
36	188	4.4	3570	2 T45025	mucin MUC5B, trach
37	186	4.4	1151	2 T18535	high molecular mas
38	184.5	4.3	1537	2 S53465	flocculation prote
39	182.5	4.3	1353	2 T19157	probable metal bin
40	182.5	4.3	1424	2 T19156	probable metal bin
41	182.5	4.3	1851	2 T19964	hypothetical prote
42	182	4.3	1075	2 S48992	flocculation prote
43	182	4.3	3869	2 A48205	All-1 protein +GRE
44	181.5	4.3	662	2 A45155	mucin FIM-C.1 - Af
45	181.5	4.3	1199	2 A40670	nuclear envelope p

ALIGNMENTS

RESULT 1

A48184 Transcription Initiation factor IID 110k chain - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C:Accession: A48184

R:Kokubo, T.; Gong, D.W.; Roeder, R.G.; Horikoshi, M.; Nakatani, Y.

A:Title: The Drosophila 110-kDa transcription factor TFIID subunit directly interacts

A:Reference number: A48184; MUID:93317591; PMID:8327460

A:Accession: A48184

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-921 <KOK>

A:Cross-references: GB:S63550; NID:9398432; PID:9398433

A:Experimental source: embryo nuclear extract

A>Note: sequence extracted from NCBI backbone (NCBI:134863, NCBI:134864)

C:Genetics:

A:Gene: FlyBase:Tafl10

A:Cross-references: FlyBase:FBgn0010280

C:Keywords: transcription Initiation

Query Match	18.2%	Score 775;	DB 2;	Length 921;
Best Local Similarity	29.6%	Pred. No. 2,16-33;		
Matches	258;	Conservative 128;	Mismatches 308;	Indels 178; Gaps 26;
QY	25	POIYAVKAPNTTIOFPANLQLPETVILIKNSGFLM-----VSPQOTVRAETTS	76	
DB	133	POSPSITLSTLNTGQTPA-----LLVKTDFGQLRVGTTGPPVOTITWTNNS	184	
QY	77	NITSAPVAPNAPNQVICTVPSNSSO-----LIKKVAVTPVKKIAOICTVTVTVTP	127	
DB	185	NITSTTNHPTTTO-IRLDTPVPAASMTNTTATSNITVSVASSGANSOPHQLTNQ	243	
QY	128	KPSVQSVAVPTSVTVTPGKPLNTVTLKPSLSLASSSTPSNEPMKAENSAVQINLSP	187	
DB	244	APQLPQIQIQTIPAQSGQQQVNVSSAGATAVSSTTA-----AFT	287	
QY	188	TMLENVK-KCKNFLMLIKLACSGSQSPDKGVKVKLVEQLIDAKIEAEFTRLKLYEIK	246	
DB	288	TOOGTKKCKRFLANLIEL--STREPRVPERKNTVTLQELVANAVEPEEFCDRLERLN	345	
QY	247	SSPOHLVPLPKSVVALROU-----LPSQSTIO---	276	
DB	346	ASPOCLLGLFKLSPLRLQALYKELVIEGKRPQHVLAGISQQLPKIAQVIRIG	405	
QY	277	---OCVOOTSQDMVIACTTIVTSPVTTVSSSQSEKSIIVSGATAPRTVS---VQT	329	
DB	406	PSQTTTIGQTVRM--TPNALGTPRPTIGHTTISKQPPN---IRLPAPRNVNNGIRT	460	
QY	330	LNPLAGVGAAGVTVLHSGPTAATGTTAGTGLQTSKPLVTSVANTVTVTSIQPKP	389	

Db 461 QIP-SLOVPGQANIVQIR--GPOHAQLOQRTGSVOIRATTPR-----PNSVPTAN----- 506
QY 390 VSGTAVTSLPAVTEGETSGAAILCPYSKPVVSPFCMDHICKPVGIGTPVOIKLAQPGPVL 449
Db 507 -----KLTAVKVGQGIKAI--TPSLHP-----PSLAISGCP-----PPTPTL 543
QY 450 SOPAGIPTGSSK---QLPSLFHYVOOPSGGNEKOVTTISHS----- 489
Db 544 SVLTLSNASTTTLPIPSLPTVHLPPALRAREOMNSLNHNSHFDKLVETIKAPSLHP 603
QY 490 -----TLTIQKCGKTMVNTIIPTSQPPASILKQITLPGKKILSLQ-----AS 534
Db 604 PHMERINASLTPIGAKTM-----ARPPAINKAIGKKKKRDAMEMDAKLTSSGGA 654
QY 535 PTOKNRKIKENVTSCEFDEDDINDVTSNAGVNLNEENACILATNSLVLTILQSKDEPFL 594
Db 655 SAANSFQOOSMSMYGDDINDVAAAGVNLAEESORILGC--TENIGTQIRSCKDEVL 713
QY 595 FIGALQKRIIDIGKKHDIITLNSDANVLISQATQERLGRLEKLTIAOHRMTTKYASEN 654
Db 714 NLPSLQARIIRAITSEAGLDEPSODVAVLISHAOERIKNTVERKLAIVAEHRIDIVIKIDPR 773
QY 655 YILCSDTRSQLKLEKIDOLEKORKDEEREMLLKAASRKNKEDDEPQULKOKAKELOQ 714
Db 774 YEPAKDVROGIKFLEELDKAEQKRHELEEREMLLRAKKSRYDEPQAKMKARAKEMOR 833
QY 715 LELAQIOHRDANTATPAIAGPRKKRPLE-----SGIEGLDNLLASGTSSTLTATKOLHRP 769
Db 834 AEHEBELRORDANTLALQAIIGPRKKRLKLDGETVSSGAGSSGGVLSGSGAPVTL---RP 889
QY 770 RIRICLRDLIFCMEQEREMKYRALLYALLK 801
Db 890 RIKRVNLRDLMFYMEQERECRSSMLFETYLK 921

RESULT 2
A:5183
TBP-associated factor TFIID - fruit fly (*Drosophila* sp.)
C:Species: *Drosophila* sp.
C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997
C/Accession: A45183
R:Hoey, T.; Wehnizler, R.O.; Gill, G.; Chen, J.L.; Dynlacht, B.D.; Tjian, R.
Cell 72, 247-260, 1993
A>Title: Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties
A:Reference number: A45183; MUID:93145326; PMID:7678780
A:Accession: A45183
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-921 <HOE>
A:Experimental source: embryo
A>Note: sequence extracted from NCBI backbone (NCBIP:123832)
C:Genetics:
A:Gene: FlyBase:Tafl10
A:Cross-references: FlyBase:FBgn0010280

Query Match 18.2%; Score 775; DB 2; Length 921;
Best Local Similarity 29.6%; Pred. No. 2,1e-33;
Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;
QY 25 POIYAVKAPNTTITOPANLQLEPPGTVLTKSNGSPLM-----VSPQQTVAETTS 76
Db 133 PQSPSTLSTLMTGOTPA-----LVKTDNGFOLLVGTGTGPTVQTITNTSNN 184
QY 77 NITSRPAVNPQTVKICIVPNSSQ-----LIKVAATVPVKLAQIGTTVTVTP 127
Db 185 NITSTNHTHTTQ--IRLQTVPAASMTNTTANSNIIVNSVASSGVANSSQPHLQLQNAQ 243
QY 128 KPSQSVAVPSPSVTVTGKPLNTYTLKPSLSGASSTPSNPNKAENSAVAQINLSP 187
Db 244 AQQLPQIQTITPAQOQOQOQNNVSSAGTATAVSSSTA-----ATT 287
QY 188 TYLENVK--KCKNFILMLIKACSGSQSPMGQNVKTVLEOLLDAKTEAEFTTKLVLEK 246

Db 288 TOGNTREKCRKFLANLIEL--STREPKVEKENVRLILOELVANNNEPEECORLERLNL 345
QY 247 SSPDHLVPLKKSVAVALROL-----LPNOSFQI--- 276
Db 346 ASQPPCLIGFLKKSILPRLRALYTKELVIGIKRPPQHVHLAGLSQQLKIQAOIRPIG 405
QY 277 ----QCVQOTSSDMVATCTTYYTTSVVTYVSSOSSEKSIYSGATAPTVS---VOT 339
Db 406 PSQTTTIGQOVNMI--TPNALGTPREPTIGHTTISKPPN---IRLPTAPRLVNTGIR 460
QY 330 LNPAGPVGAKAGVNTLHSHVPAATGCTAGGILQTSKPLVTSVANTVYVSLQPEK 389
Db 461 QIP-SLOVPGQANIVQIR--GPOHAQLOQRTGSVOIRATTPR-----PNSVPTAN----- 506
QY 390 VSGTAVTSLPAVTEGETSGAAILCPYSKPVVSPFCMDHICKPVGIGTPVOIKLAQPGPVL 449
Db 507 -----KLTAVKVGQGIKAI--TPSLHP-----PSLAISGCP-----PPTPTL 543
QY 450 SOPAGIPTGSSK---QLPSLFHYVOOPSGGNEKOVTTISHS----- 489
Db 544 SVLTLSNASTTTLPIPSLPTVHLPPALRAREOMNSLNHNSHFDKLVETIKAPSLHP 603
QY 490 -----TLTIQKCGKTMVNTIIPTSQPPASILKQITLPGKKILSLQ-----AS 534
Db 604 PHMERINASLTPIGAKTM-----ARPPAINKAIGKKKKRDAMEMDAKLTSSGGA 654
QY 535 PTOKNRKIKENVTSCEFDEDDINDVTSNAGVNLNEENACILATNSLVLTILQSKDEPFL 594
Db 655 SAANSFQOOSMSMYGDDINDVAAAGVNLAEESORILGC--TENIGTQIRSCKDEVL 713
QY 595 FIGALQKRIIDIGKKHDIITLNSDANVLISQATQERLGRLEKLTIAOHRMTTKYASEN 654
Db 714 NLPSLQARIIRAITSEAGLDEPSODVAVLISHAOERIKNTVERKLAIVAEHRIDIVIKIDPR 773
QY 655 YILCSDTRSQLKLEKIDOLEKORKDEEREMLLKAASRKNKEDDEPQULKOKAKELOQ 714
Db 774 YEPAKDVROGIKFLEELDKAEQKRHELEEREMLLRAKKSRYDEPQAKMKARAKEMOR 833
QY 715 LELAQIOHRDANTATPAIAGPRKKRPLE-----SGIEGLDNLLASGTSSTLTATKOLHRP 769
Db 834 AEHEBELRORDANTLALQAIIGPRKKRLKLDGETVSSGAGSSGGVLSGSGAPVTL---RP 889
QY 770 RIRICLRDLIFCMEQEREMKYRALLYALLK 801
Db 890 RIKRVNLRDLMFYMEQERECRSSMLFETYLK 921

RESULT 3
C87719
protein R119.6 [imported] - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C/Accession: C87719
R:anonymous, The *C. elegans* Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C-
A:Accession: C87719
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-549 <STO>
A:Cross-references: GB:chr_1; PIDN:AA016427.1; PID:g3133007; GSPDB:GNO00019; CESP:R119
C:Genetics:
A:Gene: R119.6
A:Map position: 1
Query Match 8.1%; Score 344; DB 2; Length 549;
Best Local Similarity 22.7%; Pred. No. 6.9e-11;
Matches 145; Conservative 77; Mismatches 193; Indels 224; Gaps 20;
QY 158 PSSLGASSTPSNPNKAENSAVAQINLSPYLENVKCKCKNFILMLIKACSGSQSPEN- 216

[illegible]

```

RESULT 4
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A>Title: Differential splicing-in of a proline-rich exon converts alphanac into a muscle
A:Reference number: Z20889; MOID:96312450; PMID:8696236
A:Accession: T30826
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Intons: 24/1, 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphanac into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match      5.7%   Score 244.5; DB 2; Length 2187;
Best Local Similarity    21.8%; Pred. No. ge-05;
Matches 170; Conservative 120; Mismatches 289; Indels 201; Gaps 41;

CY          2 TLVTAKAPVSAAPPKVGSRPLPAQIYAVKAENPTTT-IQFPANLQLPCTVLIKNSGRL 60
1 : | | | : | | : | :: : | || | : | | : | | : | | : | |

```

Db 1566 TIELKEADATLPSPSTKSPKIPSSK-----KAPRTSAPKEFPASPSIKPVTITSLAQTAAPS 1621

Qy 61 MLVSVQOQIVTR-----AETTSNITSR-PAVPA-----NPQYKICTVNSSQLIKKAAV 109

Db 1622 LQKASTTIPKEMNLAAAPAVLPLVSSKSPAPAPASASLSLPATAAPDTAKREAT-----TI 1675

Qy 110 TPVKKLAQIGTIVVT-----VPKPSSVSVAVPTSVVTPCKPLMTVITLTPSSIG 162

Db 1676 PSCKAAATTEPPIETSTAPSLLEGAPKETSESV---SKVLSSPPKKAS-----SSKR 1725

Qy 163 AASTPSNE-PNLKAENSAVOYNLSPTMLENVKCKNLFAMLILKAGSGOSPEEGONVK 221

Db 1726 AASTPLATPLPSLK---EASV---LSPTATSSK--DSHISPSVSDACSTGTTTPQASEK-- 1775

Qy 222 KLVEQLDPAKLEAEFTTKRLVLEKSSPOPHLVPLKKSVALROLPLNSOSGFIOCCVOQ 281

Db 1776-----LPSKCKPTAFT-----EMLAAPAP-----ESALITAPAIQKSPG----- 1809

Qy 282 TSSDMVINTCTTTVTTSPTVTTVSSSOSEKSIISGANAPRTV-----SVGTINPL 333

Db 1810-----ANSNSASSPCDDP-SSKSDTKGLPSAVALAPDTVVERKTSKAITPL--L 1857

Qy 334 AGPVAKAGVNVLLHSHVGPATATGCTTAGHGLQTSKPLVTSVANVTVTYSLQPEKPVSSG 393

Db 1858 VSPA---KGSDDLHS--FKGPVGS-----QVATPLA-----AFTSDKVPPE--ANSA 1897

Qy 394 TAVTLSPAVTFTGSETGAALICLPVSKPVVSPFCMDHICKRVIGTPOVQIKLAOPVLSOPA 453

Db 1898 SVAPKPAAPASLTLPASPVAPLPRKQPLLE-----SARGSVLESPS 1938

Qy 454 GIPPTSSSKQLESLTHVVOQPSGNEKQVTTISHSSTLTIQKCGKTKMPVNTIIPTSQPF 513

Db 1939 KLPVAEDELPEPLI-PEEAVSGGEFPQPILVN-----MPAPK--PAGTPA 1981

Qy 514 PASLIKQTLIPENKILISLOASPTOKNRJKEVNTSCFRDEDDINDVTSMAQVNLNEENACI 573

Db 1982 PAPSAAKQVPLKNNK-----GSGTESDS--DESPPE-LEEDDSIQTIATQQA----- 2024

Qy 574 LATNSELVGTLLIQSKDEBEFLFICALQKRILIDIGKRHDITELNSDAVNLISQATQERLNG 633

Db 2025 LAAAEI-----DEEP--VSKAKQSRSEKKAR--AMSKLGLHQVGTGVTATVTRKSKN 2073

Qy 634 LLEKLTALAQHMTYK--ASENYILCSPTSRQKLFLEKLDLEQKORDLERENILKRA 691

Db 2074 ILFVTV---KPDVYKSPASDTYIVFGEA-----KIELDSQQAQ-LAAAEFKVYG 2119

Qy 692 KRSKNKDEPEQRLKQKAKELQOLELAQIOHRDANLTATPAALGPPKKRPLEGIGELKDN 751

Db 2120 EAVSNIGENTQIPIYQEESEEEVDETGYEVKDIELMSQANVSAK-----AVRALKNN 2174

RESULT 5

AA0718

host cell factor C1 precursor - human

N:Alternate names: C1F; HCF; VP16 accessory protein host cell factor (VCAF)

C:Species: Homo sapiens (man)

C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 20-Apr-2000

C:Accession: AA0718; G02511; A56088; I37453

R:Wilson, A.C.; Lamarco, K.; Peterson, M.G.; Herr, W.

Cell 74, 115-125, 1993

A:Title: The VP16 accessory protein HCF is a family of polypeptides processed from a

A:Reference number: AA0718; MUID:93327419; PMID:8392914

A:Accession: AA0718

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA; protein

A:Residues: 1-2035 <WIL>

A:Cross-references: PIDN:AA827583.1; PID:9399752

A:Experimental source: Hela cell

A>Note: sequence extracted from NCBI Backbone (NCBI:135349)

R:Platzer, M.; Bauer, D.; Drescher, B.

submitted to the EMBL Data Library, March 1995

A:Reference number: H01368

A:Accession: G02511

RESULT 5
AA0718
host cell factor C1 precursor - human
N:Alternate names: CFF; HCF; VP16 accessory protein host cell factor (VCAF)
C:Species: Homo sapiens (man)
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 20-Apr-2000
C:Accession: AA0718; G02511; A56088; I37453
R:Wilson, A.C.; Lanmarco, K.; Peterson, M.G.; Herr, W.
Cell 74, 115-125, 1993
A:Title: The VP16 accessory protein HCF is a family of polypeptides processed from a
A:Reference number: AA0718; MUID:93327419; PMID:8392914
A:Accession: AA0718
A>Status: preliminary: not compared with conceptual translation
A:Molecule type: mRNA; protein
A:Residues: 1-2035 <MIL>
A:Cross-references: PIDN:AA827583.1; PID:9399752
A:Experimental source: HeLa cell
A>Note: sequence extracted from NCBI Backbone (NCBID:135349)
R:Platzter, M.; Bauer, D.; Drescher, B.
submitted to the EMBL Data Library, March 1995
A:Accession: G02511
A:Accession: G02511

A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2035 <PLA>
 A:Cross-references: EMBL:U52112; NID:91302657; PIDN:AA51751.1; PID:91302663
 R:Kistler, T.M.; Pomerantz, J.L.; Twomey, T.C.; Parent, S.A.; Sharp, P.A.
 J. Biol. Chem. 270, 4387-4394, 1995
 A:Title: The cellular C1 factor of the herpes simplex virus enhancer complex is a family
 A:Reference number: A56088; MUID:95181425; PMID:7876203
 A:Accession: A56088
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-664, 'T', 666-1637, 'E', 1639-1684, 'A', 1686-1734, 'Q', 1736-2035 <KRT>
 R:Prattini, A.; Faranda, S.; Redolfi, E.; Zucchi, I.; Villa, A.; Patrosso, M.C.; Strina,
 Genomics 23, 23-35, 1994
 A:Title: Genomic organization of the human VP16 accessory protein (HCF), a housekeeping
 A:Reference number: I37453; MUID:95130085; PMID:7829076
 A:Accession: I37453
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 100-563, 'R', 565-603, 'VS', 604-1163, 'P', 1165-1872, 'A', 1874-2035 <FRA>
 A:Cross-references: EMBL:X79198; NID:9558348; PIDN:CAA55790.1; PID:9558349
 C:Genetics
 A:Gene: GDB:HCF1; HFC1
 A:Cross-references: GDB:388714; OMIM:600196
 A:Map position: Xq28-Xq28
 A:Introns: 65/1; 114/3; 168/2; 238/1; 266/2; 302/1; 482/1; 535/3; 601/3; 676/3;
 Query Match 5.6%; Score 237; DB 2; Length 2035;
 Best Local Similarity 22.6%; Pred. No. 0.0002;
 Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;

QY 10 VSAPRVSSPRPAPQIVAVKAPNTTIOFPANLQIPGTVLKNSGRLMLVSPQOTV 69
 DB 537 IGSSPQMSGMAALAAAAATOKIPPSA---PVLVSPAGTTIVK-----MVTPEGTT 588
 QY 70 TRLETTNSINRPAVPANPOTVKICVFNSSSOLIKKVAATPVKLAQIGTV----- 122
 DB 589 LPA--IVKAVSSPVVMSNPT-----KMLKTA-----AQVTSISATNTS 628
 QY 123 ---VTVPRSSVQVAVPTSVTVTPGKPLNTVTLK--PSS--GASSTPSNPNT---- 173
 DB 629 TRPILIVHKGTV--TVAQQAQVTVVGVGVTITLVKSPISVPGSALSNIGKVMVSV 687
 QY 174 --AAENSAVQILSPMLNENKKNFL--AMLIKAGSGSSP-----EMGQNVK 221
 DB 688 QTRPVQTSATVGGASTGPVYQIIQTGKPRAGTILKLVTSADCKPTTIIITTAASGATK 747
 QY 222 KLVEQLDAKIEAEFFTRKLYELKSSPOPHLVPLKKSVAVALROLPLNSQSFLOQVQ 281
 DB 748 PTLIGI-----SSVSPSTT---KPGTTTIKTIIPMSAITTQAGATG 785
 QY 282 TSSDMVATCTTVT-----SPVVT-----TVSSSQSEKSIIVSGA-----TA 321
 DB 786 VTSSPKSPITITITTKVMTSGTGAAPAKITATPAKATGCGQGVQVVLKAPGQPGTI 845
 QY 322 PRT-----VSVQILNPLAGVYGA--AGVTLMSVGPAAATGTTAGTGLQTS 368
 DB 846 LRTVPMGVRLVTPVTSVAKKPAVTTLVVKGTVTLGTVTIVSV--SLAAGAGHS 903
 QY 369 KPLVTSVA--NNTVTSLOPEKPVSGTAVTSLPAVTGETSGAALICLPVAVPVVSCW 426
 DB 904 ASLATPITTLTGITATLSSQ-----VINPTATVSAAGTTTLAAGGLPTTITMPV----- 924
 QY 427 DHICKFVIGTPVOIKL--AOPGVLSPQA--GIPTSSSQKSLFSLFHVQOPSGGKNEKVT 484
 DB 955 -----SQPQVTLITAPSGVEAQPVHDLPPS-----ILASP-----TT 987
 QY 485 ISHSSTLTITOKCGQKTPVNTIITPSQFPAST-----LKQITLPGNKILSLQASPTQ 537
 DB 988 EOPATATVTLADSGQGVQGTIVLVCNSPCEHETGTTTATTVVAN--LGHGHPQTQ 1045
 QY 538 KNRKKNVTSQFDEDDIDVNTSMAGVNLNEKACILATNSELVGLIOSCKNEP 592

DB 1046 VQPV-----CDROEAAASLTSTVG-----QQN-----GSVVRVCSNPP 1079

RESULT 6
 A43932
 mucin 2 precursor, intestinal - human (fragments)
 N/Alternate names: mucin SMUC-41
 C/Species: Homo sapiens (hmn)
 C/Date: 10-Mar-1993 #sequence, revision 12-Apr-1996 #text, change 05-Nov-1999
 C/Accession: A49963; A45106; B45106; A43932; B3532; A61257; P00328; P00329
 R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
 J. Biol. Chem. 269, 2440-2446, 1994
 A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
 A:Reference number: A49963; MUID:94132002; PMID:8300571
 A:Accession: A49963
 A:Molecule type: mRNA
 A:Residues: 1-639 <GDU>
 A:Cross-references: GB:I21998
 R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
 J. Biol. Chem. 267, 21375-21383, 1992
 A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
 A:Reference number: A45106; MUID:93016075; PMID:1400449
 A:Accession: A45106
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 626-1895 <GUD>
 A:Cross-references: GB:I94131; NID:9186395; PIDN:AAA59163.1; PID:9186396
 A:Note: sequence extracted from NCBI backbone (NCBIP.116706)
 A:Accession: B45106
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 2037-3020 <GUD>
 A:Cross-references: GB:I94132; NID:9186397; PIDN:AAA59164.1; PID:9186398
 A:Experimental source: colon
 A:Note: sequence extracted from NCBI backbone (NCBIP.116698)
 R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
 J. Clin. Invest. 88, 1005-1013, 1991
 A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
 A:Reference number: A43932; MUID:91358717; PMID:1885763
 A:Accession: A43932
 A:Molecule type: DNA
 A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
 A:Cross-references: GB:M74027; NID:9188863; PIDN:AAA59875.1; PID:9188864
 A:Note: sequence inconsistent with the nucleotide translation
 R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
 J. Biol. Chem. 264, 6480-6487, 1989
 A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl
 A:Reference number: A43532; MUID:89197956; PMID:2703501
 A:Accession: B3532
 A:Molecule type: mRNA
 A:Residues: 1916-2193 <GUD>
 A:Cross-references: GB:M22405; NID:9188873; PIDN:AAA6334.1; PID:9188874
 A:Experimental source: intestine
 R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
 J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481; PMID:1985113
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatiri, I.; Sajjan, U.S.; McCoool, D.; Wang, D.; Jones, C.; Forstn
 Blochm, Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the
 A:Reference number: P00328; MUID:92198477; PMID:1550588
 A:Accession: P00328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M86523
 A:Experimental source: small intestine
 A:Accession: P00329

QY 415 -----LPSYKPVVSCFMDHICKPVGCTPVOIKLAQGPVLSQPAIGPTGSSSKOLESLF 468
 Db 1070 LTTIAPPSVTVTNF-----TPTITTT-----VCS-----TGNS----- 1101
 QY 469 HNVQOPSGNKKOYTTTSHSSTLTIOKCGQKTMVNTIPIPS 510
 Db 1102 -AGETSGCSKPTVTTVPCTGT-----GEYTEATTLVTTA 1138

RESULT 10

T42215
 zonadhesin - mouse
 N:Alternate names: sperm-specific membrane protein
 C:Species: Mus musculus (house mouse)
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T42215
 R:Gao, Z.; Garbers, D.L.
 J. Biol. Chem. 273, 3415-3421, 1998
 A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane protein
 A:Reference number: 22080; MUID:98123114; PMID:9452463
 A:Accession: T42215
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-5376 <GAO>
 A:Cross-references: EMBL:097068; NID:93327420; PID:93327421; PIDN:AAC26680.1
 C:Genetics:
 A:Gene: Zan
 A:Map position: 5
 C:Function:
 A:Description: functions in multiple cell adhesion processes
 A:Note: found exclusively on the apical region of the sperm head
 C:Keywords: cell adhesion

Query Match 5.0%; Score 213.5; DB 2; Length 5376;

Best Local Similarity 23.08; Pred. No. 0.014; Mismatches 260; Indels 177; Gaps 36;

QY 8 APVAP--PKVSSGPRLLPAPQIVAVKAPNTTIOFPA-----NLQUPCTVLKNSGPI 61
 Db 551 SPLRPGPSSESTVPTLMEQPTSPKATVITIEPTTPTEATITPTTTTTPTE--VI 607
 QY 62 LVSPQQTVAETTSNITSRPVANDPQVAKICIVPNSSQLIKKAVTPVKLAQIGTT 121
 Db 608 NVSKETISIPPEVT-----IPTEVTV-----SPEEIIISPTETVPPTDTAYV 652
 QY 122 VVTIV-KKPSV-QSVAVPISVTVTPGKPLNTYTIKPSL-GASSPSEPNKAE-N 177
 Db 653 EATNASEETSVPEVTLLEVTVSPEE--TVPEVPIVLIETATPEGETTLVTEVP 710
 QY 178 SAAVQINLSPTMLBNVKKCKFLMLIKLACSGSQSPMGQNVKKLVQLDAK---TEA 234
 Db 711 TVPEVGVHTEVTV-----SPE--ETSVPEETISTEVTYVSP 748
 QY 235 EETTRK-----LYELKSSQPHL-----VPFLKSVVALRQLLPNSQSFIQCVQOTS 283
 Db 749 EETTVPEVPIVLIETATSPGEITILEVTPVTEVGVHTEVTVNSP-----EETSV 802
 QY 284 -SDVAVNTCTTV-----TSPVATTVSSSQSEKSIIVSGATAPRTVSVOQLNPLAGV 338
 Db 803 PTEETISTEVTVPPEETTLTEVPTVSTEVTVS-----PEETSVPEETI----- 849
 QY 339 AKAGVTLHSVGPATAGTGTAGTL-----QTSKPLVTSVANTVTVS-----LOEK 388
 Db 850 -----LTLITLEVPIVPEVTEVGVHTEVTVNSPEETISTEVTVPPEETTLPTREV 905
 QY 389 PVSGVAVTISLPAVTEGEGSGAICLPVAK-----VVSFCMDHICKPVIGTVOIK 441
 Db 906 PTVS-TEVTVNSPEET-----SVPEETIILEITVSPPEETVPIEGTTL----- 949
 QY 442 LAQGPVLSQPAIGPTGSSSKOLESLFHVQOPSGNKKOYTTTSHSST----- 490
 Db 950 -----PTEVLIIVPIEVTTPGTGT--VPTEVPTVSTEMGTGVHTEVTVPEETISTEVTAT 1004

QY 491 -----LTIQKCGQKTMFVN-TIPTSQPP--ASILKQITLPPGNKILS 530
 Db 1005 VLPASIPPEETTTTPEVTTTTPPEETTPAEVTVVPAIPPEETIASLEVTTPPEETTT 1064
 QY 531 ---LQASPTOKNRKENT-----SCFDEDDI-NDVTSMAQ--VNLINEACILATNSE 579
 Db 1065 PTEVTVPEETTTTTPTEVTVVPAIPPEETTVPEETTVASSETTVSTOETTLTLEQSA 1124
 QY 580 LVGTLL---QSCKDEPFLFICAL 599
 Db 1125 VTQTSIACRPPCPSPPLMIPPL 1147

RESULT 11

T34434
 hypothetical protein K06A9.1a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T34434
 R:Geisler, C.; Gattung, S.
 submitted to the EMBL Data Library, December 1996
 A:Description: The sequence of C. elegans cosmid K06A9.
 A:Reference number: 221525
 A:Accession: T34434
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2232 <GEI>
 A:Cross-references: EMBL:080846; PIDN:AC70890.1; GSPDB:GN00028; CESP:K06A9.1a
 A:Experimental source: strain Bristol N2; clone K06A9
 C:Genetics:
 A:Gene: CESP:K06A9.1a
 A:Map position: X
 A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20

Query Match 5.0%; Score 212.5; DB 2; Length 2232;

Best Local Similarity 21.9%; Pred. No. 0.0046; Mismatches 138; Conservative 73; Mismatches 247; Indels 173; Gaps 24;

QY 26 QIVAVKAPNTTIOFPALQLPCTGVILKNSGPIMLVSPQOIVRAETTSNITSRAVP 85
 Db 801 QSVSTNSPGSTVTR-----PSTVSGSTSSGTVTVGSTEASTSGSVASSSPAPSTS 852
 QY 86 ANPQTVKICIVPNSSSQLIKKAVVPAVKLAQIGTVTVTPKPSVQSVAVPTSVTVT 145
 Db 853 QNPV-----PSTNS-----GSMITQSPYRS--QSTSPVESSTIPS 886
 QY 146 PKPLNVTTLKPS-----SLGASSTPSN-----EPNL 173
 Db 887 PGSPPTTLTSTSPSPSQSTTIGSTGSPGISTSEMTSGSGSTQTPGSGTVTQPSST 946
 QY 174 KAENSA--VOINS-----PTMLENVKCKKFLMLIKLACSGS-----QSPKMON 219
 Db 947 VSDSTSSGTVVSGTEGSSSPIPSTQNTNPSTV-----SGSSMSTQTPQSSQS 996
 QY 220 VKLVEQLDAKIEAEFTTKLYELKSSQPHL-----VPFL 257
 Db 997 TSP-VESTSGATSSSGSPGTTLSISPSRPSSTIGSSQGSTSVYSTISQGSTETPGS 1055
 QY 258 KKSVALRQL-----PN-SQSFIOQCVQOOTSMDVIA 289
 Db 1056 TGSFTVKPSTVSGSASGSTATMGSTEASTSGSGSTSPNSQSTSPSTSGATSSPGSSG 1115
 QY 290 TCTTTVTPSVVTTVSSSQSEKSIIVSGATAPRTVSQVQLNPLAPGAKAGVTVLHSV 349
 Db 1116 TTLVTSIPSPSQSSTIGSSQSTSPVSTSGDMTSQGSTQIP-----GSGSTVTVQPSST 1170
 QY 350 GPATAT-----GTTAGTGLQTSKPLVTSVANTVTVVSLQPEKPVSGTAV---TL 398
 Db 1171 GSGSTSTGELTISQGSTOTPPSSLSSTPAISTSTQOASTNS--PGSTVTVQPSVTRGSTS 1228
 QY 399 SLPAVTEGETSG-----AAICLPVAVVSCMD-HICKPVIGTVPVOIKLAQGPVLS 450

Db 1229 SGGTWTGTEGSSSTGSSSATSLSSSSPVPSSTQSPNPSTGSSSTPTPNPQSTSPVVS 1288
 Oy 451 QPAGITGSSSKQLFS-LRHVVOQP---SGGNKQVTTTSHSSTLTIOCKGQATMPVN-T 505
 Db 1289 TTTGEMTSHGSTCTPTSTISTYQTPSTVSGSNSGSTVIGSSEASTSGSSFTSPSSIS 1348
 Oy 506 IIPTSQPPASILKQITLPGNKILSLQASPT 536
 Db 1349 PVPSTSPISPTFASST-SGSTISDVSSVST 1378

RESULT 12

t34369
 hypothetical protein t19d12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34369
 R:Javello, A.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid t19d12.
 A:Reference number: 221513
 A:Accession: T34369
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1777 <FAV>
 A:Cross-references: EMBL:U41263; PIDN:AAC24428.1; GSPDB:GN00020; CESP:T19D12.1
 A:Experimental source: strain Bristol N2; clone t19d12
 A:Gene: CESP:T19D12.1
 A:Map position: 2
 A:introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681

Query Match 4.9%; Score 210.5; DB 2; Length 1777;
 Best Local Similarity 22.1%; Pred. No. 0.0043;
 Matches 162; Conservative 100; Mismatches 266; Indels 205; Gaps 38;

Oy 2 TLVTKVAPASAPPKVSSGRLPAPQIV-AVKANNTTIOFPANLQLPQGVLIKNSGRL 60
 Db 453 TILQSPTPTSTPLVSSSSGSSSTVYTSTITSTOGVPTSTNOPTPT-----SNP- 505
 Oy 61 MLVSPQOITVRAETTSNITSRPAVPANPOTVKICTVPS-----SSQLIKKVAVTPYKLL 115
 Db 506 -TTPKSVTASPTSGANS---TASPTSTI-TSAPTSQSHSPSSSTMTSTVPT--STP 556
 Oy 116 AQTGTIV-VTVPKPSSVQS-VAVPTSVTVT-----PKRPL----- 150
 Db 557 ASTTTTVPITVAPGQCYCQSNVAFAELTSGTSDLDLDIONFIANTLFFYSAGAPYLLGL 616
 Oy 151 ---NVTTLKP-----SSIGASSTPS-----NEPNLKAENSAVQ--INTSPM 189
 Db 617 TONRTATISLVPPNDOTLSDMTYGAEQTPSGISALDLTFNILARGNAVISDAFNITP-- 674
 Oy 190 LENVKCKNFLMLIKLACSGSOSPENGVNKKVLEQLDA--KIDAEETRLKLYELKS 247
 Db 675 --NTR-KGYQGFVLVA-----NSDESQASVDSATNLKAQGF-NVITVAFKS 719
 Oy 248 S-----PQPH---LVFLKKSVA-----LRQLLPNS-----QSFIOQ 277
 Db 720 SKKFDVLASQPSNYNTIYODADKRYVATLIGNVLTNNLNLAASTMTSGVYSSTVY 779
 Oy 278 CUGQOTSSDVAIACI--TVVTTSPVYTTT--VSSSQ-----SEKIIYSGAT-APRT--VS 326
 Db 780 TTQOGSSQAPSSIVYPTTGTSIGAASITGISTIOQANSTSSVITTGTSAPQSSSTANS 839
 Oy 327 VQTLNPLAGPVAKAGVTLHSGVPTAATGTTAGTLOTSKPLVTSVANTVTVSLQ- 385
 Db 840 STTTSB-----STTGSTPAPOSSTVASTTVSPYTTTECICITVSNFTGCTSTTG 890
 Oy 386 -----PEKPVSGTAVTTLPLAVTBETSGAICLPVSPVVSFCMDHICKPVIGTPVQI 440
 Db 891 LTSSASQSTASGVSSTVASTTIPQSSS-----SSPOSP 925
 Oy 441 KTAQGPVLSPAGIPTGSSSKQLFSLFHVVOQPSGCKNEKQVTTTSHSSTLTIOCKG--- 497

Db 926 TSQAPQSSSTSAATYVSSS-----QSFSTSPAQSSSTPAQSSSTVYQSSSFQ 975
 Oy 498 -QKTMV--NTIIPSOFPASILKQITLPGNKILSLQASPTOKNIRKNVTSCEFDED 553
 Db 976 SPOSTQIGSSTVTPSQVASSST-----SGGPTTQICPNQGVFKQGVYI---E 1024
 Oy 554 DINDVYSMAQVNLNEEACILATNSLVGLIQSCADEPLFGAL-----QKRIIDIG 607
 Db 1025 MLPASTQQNAIMAFEVENVLL--NSMEYGLALDNLTHDNRTLVTAIYPPTDYKNVQYXG 1081
 Oy 608 KKHDTLENSDAV 620
 Db 1082 SANSVDEFKQVI 1094

RESULT 13

A47283
 calphotin - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster
 C:Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000
 C:Accession: A47283
 R:Ballinger, D.G.; Xue, N.; Harshman, K.D.
 Proc. Natl. Acad. Sci. U.S.A. 90, 1536-1540, 1993
 A:Title: A Drosophila photoreceptor cell-specific protein, calphotin, binds calcium
 A:Reference number: A47283; MIMD:93165730; PMID:8434015
 A:Contents: photoreceptor cells
 A:Accession: A47283
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-873 <BAL>
 A:Cross-references: GB:105080; NID:g157071; PIDN:AAA28420.1; PID:g157072
 A:Note: sequence extracted from NCBI backbone (NCBI:124958, NCBI:124959)
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 4.9%; Score 210; DB 2; Length 873;
 Best Local Similarity 21.5%; Pred. No. 0.0017;
 Matches 118; Conservative 82; Mismatches 239; Indels 110; Gaps 22;

Oy 3 LVTKVAPASAPPKVSSGRLPAPQIVAVKAPNNTTIOFPANLQLPQGVLIKNSGRL 62
 Db 29 VVSPAAVAVAPAVAPAPAPAPAVTVPAPPTLASVQ-PATVAVPAPAPAPAAVAVAS 87
 Oy 63 VSPQOITVRAETTSNITSRPAVPANPOTVKICTVPS-----SSQLIKKVAVTPYKLL 116
 Db 88 VAP--PVVAPPTP-----AASPVSTPPVAVAOIPVAVSAPVAVPAPVAPPTPVAPVPA 140
 Oy 117 QIGTT--VVTVPKPSVQSAVAPTSVTVTPCKPLMTVTTLKPSISGASSTPNEPNK 174
 Db 141 PVATPPVAAAPTPAATPVVSP--VIATPPVVPANTV--PVAPVAAVAPVAVPVA 195
 Oy 175 AENSAVQINLSPTMLE-----NVKCKNFLMLIKLACSGSOSPENGVN 221
 Db 196 PVLAAPAVAPAVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 243
 Oy 222 KLVEQLDAKIEAEETRLKLYELKSSPOPHLYPFLKSSVALROLNPSQSTIOQVOO 281
 Db 244 KPLAAEPVVPVAPATETPVAPAAASPHVSVAPVAPVAPVAPVAPVAPVAPVAPVAP 290
 Oy 282 TSSDMVATCTTCTVTTSPVTVTVSSQSEKSIIVSATAPRIVSVQTLNPLAGPVAKA 341
 Db 291 TEPVAAATITTPER-PALAPVYASQVAAVAVVATPPPP-APEETIAP----- 339
 Oy 342 GVVTLHSGVPTAATGTTAGTGLQTSKPLVTSVANTVTVTVSLQPERPVSGTAVTSLP 401
 Db 340 -----PVVAETPEVASVAVAETTPPVVPA-----ASIDAPVATITPVATIL- 383
 Oy 402 AVTFGETSGAICLPVSKPVVS-----FCMDHICKPVIGTPVQIKLAQGPV 448
 Db 384 AVTDPPVTAASV--PELIPVAPSPVPSVAETRPVVDLAPVLPVAPVAPVAPVAPVAP 441
 Oy 449 LSPAGIPTGSSSKQLFSLFHVVOQPSGCKNEKQVTTTSHSSTLTIOCKGKQKMPVNTIIP 508

Db 442 TPAPASPVITIALDIPVAVIAPSDAPAE-----APSAAPI---VSTPPTTASVP 492
OY 509 TSOPFPASI 517
Db 493 ETAPPAV 501

RESULT 14

A47282
calcium-binding protein calphotin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A47282
R:Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993
A:Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
A:Reference number: A47282; M01D:93165729; PMID:8094559
A:Accession: A47282
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-865 <MAR>
A:Cross-references: GB:L02111; NID:9157031; PIDN:AAA28405.1; PID:9157032
A:Experimental source: photoreceptor cells
C:Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBIIP:124956)
C:Genetics:
A:Gene: FlyBase:Cpn
A:Cross-references: FlyBase:Fgn0010218
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: calcium binding

Query Match 4.8%; Score 206.5; DB 2; Length 865;
Best Local Similarity 21.7%; Pred. No. 0.0025;
Matches 122; Conservative 82; Mismatches 242; Indels 115; Gaps 22;

OY 1 GTLVTKV-APYSAP-----PKVSSGRRLAP---QIVAVKAPNTTIOFPALQ 45
Db 4 GTTSPASAPVAPVTSVSAVAPVQVSPAAPVAPVAPVAPVAPVAPVAPVAPVAPV 63
OY 46 LPPTVLKSNQSGPLMLVSPQVTVRAETTSNITSRPVPANPQTVKICTVPNS--SSOLI 104
Db 64 IPAPAPIAASVTPVAVASAP--PVVAPTP-----PAASVSTPVPVAVQIPVAVASAVA 115
OY 105 KKAIVTPVKIL-AOIGTTVTTVPKPSVSVASVAVPTSVT--VTGPKPLNTVTLKPSLSG 162
Db 116 PPAVATPTPVQIPVAVATPVPAASAPVPAVTPVISPVIASPPVVPANTVPAVAP 175
OY 163 ASSRPSNEPNLKAENSAVQVNLSPMLE-----NKKCKNPLMLIKLACS 209
Db 176 VAAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAVPAV 227
OY 210 GSQSPMGQNVKRLVEOLLDAKIEAEFTKRLVELKSSQPPLVPLFKSVVALROLLP 269
Db 228 ----PEVSVATKRLAAAEFVAVAPRATETPVVAPAAASPHVAVAPVETAVVA-----P 278
OY 270 NSQSFIOCCVOQTSDDVIATCTTTVTTSVVTTTSSQSEKSIIVSGATAPRTVSQ 329
Db 279 VSAS-----TEPPVAAATLTAPET--PALAPVAVESQVAAVAVVATPTPT--APEDET 328
OY 330 LNPAGVGAAGAVVTLHSVGTATAGTGTAGTGLQTSKRLVTSVANTVTVTSIQPKP 389
Db 329 IAP-----PVVAETPEVASVAVAEETTPVPVPA-----AESIPAP 364
OY 390 VSGTAATLSPAVTFTGSGAALCLPSVKPVVS-----FCMDHICKPVIGT 436
Db 365 VVATTPVPAITL-AVTDDEVTAASV--PELPVVIAPSPVSAVAETPVVLAAPVPAVAAE 421
OY 437 PVQIKLAQPGFVLSPAGICTPGSSSKOLFSLFHVVOQPSGNGEKQVTTISHSTLLTKC 496
Db 422 PVPAVVAEEETPEPAPASAPVITIALDIPVAVIAPSDAPAE-----APSAAPI--- 473
OY 497 GQKMPVNTIITPISQFPASI 517
Db 474 -VSTPPTTASVETTPAPPAV 493

RESULT 15

T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34433
R:Geisler, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34433
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1032 <GEI>
A:Cross-references: EMBL:U08046; PIDN:ACT0889.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2;

Query Match 4.8%; Score 206; DB 2; Length 1032;
Best Local Similarity 22.5%; Pred. No. 0.0034;
Matches 141; Conservative 79; Mismatches 246; Indels 160; Gaps 26;

OY 2 TLVTKVAVSAPPKVSSGRRLAP---QIVAVKAPNTTIOFPALQLPQGVILKSNQ 58
Db 246 SISTRALPI-ASSSASSPSAASSTTPVLLSSTIQSSSGTFPSSVASSPSTV--GSTG 302
OY 59 PLMLVSPQVTVRAETT--SNITSRPVPANPQTVKICTVPNS--SSOLI 117
Db 303 AASSSVATVSTIAGTSTSTT---PVGSSSTISSSTPSSASSSSSGTMTS-----GS 354
OY 118 IGTVVTTVPKPS-----VQVAVPTSVTVTPGKPLMTVTLKPSLIGASSTPSNE 170
Db 355 TGST-VTVVPCSSSTFPASTPPLASSSSPGSTVPAAGS-----SSTVGSSTPS-- 401
OY 171 PNLKRNAAVQINSPMLNLEVKCKNPLMLIKLACSGSPMGQNVKRLVEOLLDA 230
Db 402 -----ASSSSSGTMTSNGSTGTVA 424
OY 231 KIEAEFTKRLVELKSSQPPLVPLFKSVVALROLLPNSQSFIOCCVOQTSDDVIAT 290
Db 425 PVSSSTFG-----SSTP-----IASSSSSGTTVVSSSSSTTSGS 460
OY 291 CTTVTTSPVTVTTVSSQSEKSIIVSGATAPRTVSQTLNP-----LAGPVGARAG 342
Db 461 TPSASSSSAGTASTISGSGTATVPGSSSSVGSSTQASAPSPGTMSTVSGPTGSTVT 520
OY 343 VVTLHSVPT-----AANGGTAGTGLQTSKRLVTSVANTVTVTSIQPKPV 391
Db 521 VVPGSSTPAPSSSPNPSSTPASTGTSTTISG---SSIIIVTSGS-----TV 566
OY 392 SGTAVTTLSPAVTFTGSGAALCLP-SVKPVVSFCMDHICKPVIG--TPVQIKLAQPGV 448
Db 567 SGTSTISOSTLASSTATGSSSTVPSSSPQS---SOPAPNPGTSTPSTSSSPSPS 623
OY 449 LSQAGIPTGSSSKOLFSLFHVVOQPSGNGEKQVTTISHSTLT--OKGQKMPVNTI 506
Db 624 MNPSSSTPTGSSQSTITPEGSTVASSPTG-----STGSTFVATVTVTQSTVPSGSS 674
OY 507 I---PTSQFPASITLKOTILPKNLISLQASP--TQKRIENVTSCREDDEDIDVATM 561
Db 675 LGTOSTNSPSPSSLSPT--SGMSTLTSEPSSTQSSGAOSTLT-----PSPNPQST 728
OY 562 AGVNLNENACLIATNSELVGLTLOS 587
Db 729 SLESSTSGA---TTSSSGAGTMTS 751

Search completed: February 16, 2003, 22:00:16

Thu Feb 20 16:36:13 2003

Job time : 51.8257 secs

us-09-763-909-2.rpr

Page 10


```

Db 108 TTVQPTTTTQDRSPGVQPOLVIGGSAQPAISLTATAVOTGTPOKTVPGASTTSTAANE---164
Oy 186 SPMLENVKKCKNFILMLIKLACSGSOSPEMGONKKLYEQLLDIAKEAEFTKRLVEL 245
Db 165 ---TMENVKCKSEFSLTLIKLASSGOSTETANVDLYONLDGKIEADEFTSRLEYREL 221
Oy 246 KSSPOHVLPELKKSVAVLROLLPNSOSFIOCVQOOSDMVATCTTTTSPVYTVV 305
Db 222 NSSPOPLVPLFKSLPALQSLTPDSNAFIQSSQOQPP---ASQATTLATVAVLSSSV 277
Oy 306 SSSQSEKSIIVSGATAPRVYVOTLNLPLAGPVAKAGVYVLSVGPATAGTTAGTGL 365
Db 278 ORTAGTAASVTSALQPVVLSL-----TQPVGVGKQ 310
Oy 366 QTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPAVTFGETSGAICLPSKPVVVSFC 425
Db 311 APPPLVIO-----QPPKP---GALRPQVTVLTQT-----PVALR 344
Oy 426 WDICKPVIGTPVQIKLAQGPVLSQPAIGTSSSSKQLSLFHVVOQPSGNEKQVTTI 485
Db 345 QPH-NRIMLTTPQOQL-----360
Oy 486 SHSSTLTIOKCGQKTPVNTIIPTSQFPASILKQITLPGNKIL---SLQASPTOKNRK 542
Db 361 -----NOLQPVVVKPVYLPETKALSTVSAQAAAAQKNKKK 396
Oy 543 ENVTSCFRDEDDINDVTSAGVNLNEBNACILATNSLVGTLIQSCDEPFLGALQKR 602
Db 397 EPGGSGFRDDDDINDVASMAGVNLSESAIILATNSLVGTLTRSCDDDFLLPAPLQRS 456
Oy 603 IIDIGKKHDTLNSDAVNLISOATQOERLGLLEKLTALQOHRMTYKASENTYILSDTR 662
Db 457 TLEIGKKGHTDCHPVDVSVSHATQORLQNVKEKISERQAKNFSTKDDDRYEQASDVR 516
Oy 663 SOLFLEKLDQLEKORLDEEREMILKAASRSNKEDPEQLRLKOKAKELQOLELAQIHR 722
Db 517 AQLKFFQLOLQIKORRDEQERELMRAKRSRQEDPEQLRKQKTKEMQOQELQAMQR 576
Oy 723 RQANLTATATAGRRKRPLP-----SGIEGLKNLNLASGSSLTATQOLARPRITRCLR 777
Db 577 RQANLTATATAGRRKRKYDCTGTGSGAOGSGPAAVPGSGVGTPOFROTTRITRVNLR 636
Oy 778 DLIFCMEQEREMKYSRALYLALL 800
Db 637 DLIFCLENERETSHLLLYKAF 659

RESULT 2
OBT9EO PRELIMINARY: PRT: 851 AA.
AC 0819E0:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE SD04735P.
GN TAF110.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Broksstein P., Hong L., Agbayan A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacieb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Xu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY069807; AAL39952.1; -
SQ SEQUENCE 851 AA; 92093 MW; 665B28B9588C984C CRC64;

```

Query Match 17.7%; Score 753.5; DB 5; Length 851;

```

Best Local Similarity 30.5%; Pred. No.1.6e-34;
Matches 263; Conservative 122; Mismatches 291; Indels 187; Gaps 31;

Oy 22 LPAQIVAVK--APNTTIO--FPANLQLPQGTGLIKNSGPMILVSPQOTVTRAEYSNI 78
Db 93 LPAQVYVGMHQAPPSQOQKAMPTN---PLSRVYNSHMGAVRQSP--SMTYNTATANSI 148
Oy 79 TSRPAPVANDQYKICTVPPSSS--QLIKRVATVPYKLAQIGTVTVTVPKRSSVOSA 136
Db 149 I-----VNSVASSGVANSQPPHLLQNLNAP--QLPQI--TQIQTIPAOOSQO--- 193
Oy 137 VPTSVVTVTPGKPLNTMTTLKPSLSGSSVPSNEPNLKAANSAAVQINLSPTMLEANK-K 195
Db 194 -----QOYNNVSSAGTATVASTTA-----ATTQOGNTKEK 226
Oy 196 CKNFELMLIKLACSGSOSPEMGONKKLYEQLLDIAKEAEFTKRLYVELKSSPQHLVP 255
Db 227 CKRFLANLIEL--STRPKRVKKNVRLIQLVNAVNEPPEFDRLERLNASPQOLIG 284
Oy 256 FLKSSVALRQL-----LPSQSPTIQ-----OCVQO 281
Db 285 FLKKSPLRLQALYTKELYEGIKPPQHVLAGLSQQLPKIOAQIRPIGSGQTTIGQ 344
Oy 282 TSSDMVATCTTTVTSVTVTVTVSSQSEKSIIVSGATAPRVSS--VQTLNPLAGPVG 338
Db 345 TQVRMT--TNNALGTPRPTIGHTTISKQPN---IRLTPAPRVNLTGIGITQIP-SLOVP 398
Oy 339 AKAGVYVLSVGPATAGTTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTL 398
Db 399 GQANIYQIR--GPHQHLQHTGSGVQIRATRP-----PNSVPPAN-----456
Oy 399 SLPAVTFGETSGAICLPSKPVVSGFWDHICKPVIGTPVQIKLAQGPVLSQPAIGT 458
Db 437 KLTAVKVGQTOIKAI--TPSLHP-----PSLAISGSP-----PPTPLVSLSTLSA 482
Oy 459 SSSK--QLPSLPHVVOQPSGNEKQVTTSHS-----TLT 492
Db 483 STTTLPLPSLPYVHLPEALRAREQMONSLNHSNHFDAKLVEIKAPSLHPMERIMAS 542
Oy 493 IORCGQKTMPEVNTIIPTSQFPASILKQITLPGNKILSLQ-----ASPTOKNRKE 543
Db 543 LTPIGAKTM-----APPALNKALIKKKRDMEMDAKLNITSSGGAASANSRPFQO 593
Oy 544 NVTSCFRDEDDINDVTSAGVNLNEBNACILATNSLVGTLIQSCDEPFLGALQKRI 603
Db 594 SSMSSMGDDDDINDVAMAGVNLAEESQRIIGC--TENIGTQIRSCXDEVFLNLPISQARI 652
Oy 604 LDIGKKHDTLNSDAVNLISOATQOERLGLLEKLTALQOHRMTYKASENTYILCDTRS 663
Db 653 RAITSEAGLDEPSODYAVLISHACQERLKNIVERKLAVIAHRIDVITKLDPRYPADVRG 712
Oy 664 QLFLEKLDQLEKORLDEEREMILKAASRSNKEDPEQLRLKOKAKELQOLELAQIHR 723
Db 713 QIFLELELDAEQKRHELEEREMILKAASRSNVEDPEQAKMKARAKEMQRAEMELROR 772
Oy 724 DANLTATATAGRRKRPLP-----SGIEGLKNLNLASGSSLTATQOLARPRITRCLR 778
Db 773 DANLTATATAGRRKRLKLDDETYSGSSGSSGVVLSGSGAPPTL---RPRIKRVNLRD 828
Oy 779 LIFCMEQEREMKYSRALYLALL 801
Db 829 MIFYMEQEREFRCSSMLFKTYLK 851

RESULT 3
OBT9EO PRELIMINARY: PRT: 549 AA.
AC 061707:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Hypothetical 60.4 kDa protein.
GN R119.6.

```



```

Db 2126 PAKAQTATQAPKARPPQPOPOSPAPPEVOTQTTVSSHVSEAKPTIA----- 2179
Oy 361 GTGLAQTSKPLVTSVANVTVSLQPEKPVSGTAATLSLPAVFGETSQAALCPSPVR 420
Db 2180 -----QSKRPQVAAGS-----QPSNVQGSPPVQSPSQTIRPSPQSLSPGQGS 2226
Oy 421 VVSFCDHICKPVIIGTPVQIKLAQPPVLSQPPGIPPTSSSKOLFELHVVQOPSGGNER 480
Db 2227 QVO-----TTTSTQPI---PIQPHSTSLQIPSOQOPQSOPOVOSSTQTL-----SCOTLN 2272
Oy 481 QVTTTSHS-STLTIOKCGQKTPMVTIIPTSQF--PPASIKOTLTGKNILSLQASPTOK 538
Db 2273 QVSVSSPSPQLOIQ-----PQPVIAVPQLOQOVVLSQI---QSOVVAQIQ-----QO 2321
Oy 539 NRIKENV-----TSCFRDEDDINDVTSNAGVNLNEENACILATNSLVGTLLQOSCK 589
Db 2322 SCVPOQIKQLPQLQOASAVQTHQIONVTVQAASVQOL-----QVQOQLR 2369
Oy 590 DEPFLEIGALQK---ILDIGKKHDTLELNSDAVNLISQATQERLGLLEKLTAKQH-- 644
Db 2370 DQO-----QKKQOQIIEKHETLQASNO-----SEITKO---VVMKHAVALTEHAK 2413
Oy 645 ---RMTTYKASEN--YILCDSRSLQFLEKLDQLEKORRDLERMLTKAKSRGKED 699
Db 2414 QKSMTPARERENQNVICNVQMKYI--LDKIDKEKO-----AAKKRREES 2459
Oy 700 PEQRLKQAKELQOL 715
Db 2460 VEQKRSKQNAKTLSAL 2475

```

RESULT 5

```

OQ8TE50 PRELIMINARY; PRT; 1322 AA.
AC Q8TE50;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE EMST protein.
GN C11ORF30.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RA SEQUENCE FROM N.A.
RA Hughes-Davies L.;
RT "EMST is amplified in breast cancer and displays a BRCA2 dependent DNA
RT damage response";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430203; CAD22881.1; -.
SQ SEQUENCE 1322 AA; 141467 MW; 7F8G95EBBA0FC9F0 CRC64;

```

Query Match 5.9%; Score 250.5; DB 4; Length 1322;
 Best Local Similarity 20.0%; Pred. No. 7.4e-06;
 Matches 207; Conservative 157; Mismatches 370; Indels 301; Gaps 48;

```

Oy 5 TVVAPVADPKVSSGRPLPAPQIVAAKAPNTTIOFPAN--LQLPG--TVLTKSNSGLPM 61
Db 149 TTSFTSTP-----VPSGIATVKSPPRAS--PASNVVVLPGSGTVYVKSVCSDPE 197
Oy 62 LVSPOQVYRAETSNITSRPAVPANPQV---KICVPPNS---SOLIKVA----- 108
Db 198 DEKPK---RRRTNSSSSSVVLKEVPKAVVPSKTIITPVGSGPKMSIMOSIASLPP 254
Oy 109 -VTPVK-----KLAQIGT---VVTVPKPSVQSV-----AVPNSVTVV 144
Db 255 HNSPVKTIPTKSTQTTTQKVIITVSPSSTFVPNLSKSHNVAATKLVPISVVAS 314
Oy 145 TPGRKPLNTVTLTKPSL-----GASSTPSNEPN-----LKAENSA 180
Db 315 TQGRP-PVYITASQSLVSSSSSSSSSPSPIPNVTAVVSSTPSVVMTVAQGVST 373

```

```

Oy 181 VQINLSPMLNENKCKKNFLAMLIKLAGSGSPENQGVNKKLVEOLLAKTIE--EETR 239
Db 374 SAIKMASTRLPSPKSLVSAPTQIL-----ADPPKHQOOSPK--QOLYVQOQTFQOQVAA 425
Oy 240 KLVLELKSSPRLHVLPEFKRSYVALRQLPNSQSFQOCVQOQSSDMVATCT-----YV 295
Db 426 PSEVSHQOQDOSPDLPGIKPTIQIKO-----ESGAKITIQOVQPSKILKPVYATLPTS 480
Oy 296 TTPSVVTVTVSSQSEKSIIVSGATAPRVSVQTLNPLAPGAKGVVY----- 345
Db 481 SNRPVIVSSNGAIMTKLVTTPTGQATYTRTVPSPISGMATPGATYVKTGSGII 540
Oy 346 -----LHVG-----PRTAGGTAGTGLQTSKPLVTSVANTV---TVSLQPEKPV 391
Db 541 TVVPKSLATLGGKIISNINVSQTTTKITITIPMSKPNVIVQKTTGKGTIOGLPEKNAV 600
Oy 392 S-----GTVATLSLPAVFGETSQAALCPSPKPVVSCMDHICPV--IGTPVQIK--- 441
Db 601 TILLNAGEKTIQTV-----TGAKPAILTATRPITKMI---VTPKGIQSTVQPAKI 651
Oy 442 -----LAQGPVLSQAPGIPGTSSKQLFSLFHVQOPS----- 475
Db 652 IPTKIYGGQKQTVLIKRPVTFQATV--SQGTQVLT--ETLQASRVNLAGSSSTQ 707
Oy 476 -GGNEQVTTISHSTLTIOKCGQKTPMVTIIP-----SOPPASILKOTL 523
Db 708 EGKEEPQNTYDSSSTESSQSDSQPVHVYASRQDSEHEIAMETSPITTYQDVSS 767
Oy 524 PGKILSLQAS--PQOKNRKENVTSQFRDEDDINDVTSNA--GVNLNE-----ENACIL 574
Db 768 ESQASATIKALLELOQTTVKELSS--KPROPTDLSQNAVPIQMTQKRHSPESSIA 825
Oy 575 ATNSELVGTLLQSCKDEPFLIGALOKRIIDIGKKHDTLELNSDAVN-----LISQATO 628
Db 826 VYSEVELAEVI-----TTERDEGEVAFPLVLSRSP 859
Oy 629 ---ERLGLLEKLTIAQHRMTTYKA-----SENYILCDSRSLQFLEKID 672
Db 860 OQPSQPORLLOH--VAOSQFATQTSVVVKSIPASSPAGITIHMOQALSHMTAFKHS 916
Oy 673 QLEKORRDLEREML-----LKAASRKNEDP--EQRL-----KQKA 709
Db 917 ELTGECEVEMDTPDPQGLFRSALTSQSSQKQKLSQPPLEQTOGLVKTLQCFQTKQ 976
Oy 710 KELQLELAQIQHRDANLATAIGPRKKRPLESIGLEKDNILASGTSLLTATKQLHRP 769
Db 977 KQTHIQANQLOKHLQPM--PQLSIRHQKLTPLQOEOAQKRPV-----QHTQHP 1024
Oy 770 RITRICLRDLIFQMEQREKYSRALYLALKZPLHSHIHILAIYCORRHKALLHC--P 826
Db 1025 MVAK-----DROLP-----TLMAQPPQTVQVLAVKYTKQQLPKLQQAQNP 1065
Oy 827 EISISGKZHOHERAL 841
Db 1066 KTIYQPTQPOSQMSL 1080

```

RESULT 6

```

P70670 PRELIMINARY; PRT; 2187 AA.
AC P70670;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE NASCENT polypeptide-associated complex alpha polypeptide (Alpha-NAC,
DE muscle-specific form GP220).
GN NACA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]

```



```

OY 512 FPPASI-----LKQITLPGNKILSLASPTOKRIKENVTSCFDEEDINDVTSMAGV 564
DB 949 NPPCFHERGTNTATTATTYVAN--LGSHQPTIOYOFV-----CDQOETASLVTSAVG- 999
OY 565 NUNEACILATINSELVGLIOSCKDEP 592
DB 1000 ---OQN-----GNVRYCSNMP 1013

RESULT 8
OQFMHO
AC Q9FMHO PRELIMINARY; PRT; 689 AA.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE GblAAR24960.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones.";
RL DNA Res. 4:401-414(1997);
DR EMBL; AB008267; BAB08277.1; -.
SQ SEQUENCE 689 AA; 75680 MW; 4A028C47111AFA3E CRC64;

Query Match 5.5%; Score 234; DB 10; Length 689;
Best Local Similarity 19.7%; Pred. No. 2.7e-05;
Matches 143; Conservative 125; Mismatches 258; Indels 200; Gaps 29;

OY 167 PSNERPLKAKENSAVOI---NLSPTMLENVKCKCNFLAMLIKACSGSQSPENGVNKKL 223
DB 64 PVNPMVNPINRNPKOVFPFALLPTLMNQLDK-----DRALQLRTL 104
OY 224 VQOLIDAKLEAEETRKLVLELKSSQPHLVLPFKKSVALQQLPNSQSFIOCCYQQRIS 283
DB 105 YARKLKNEIPKEGFTRHM-----KDIVG-DQMKMAVSKIQQ-VVYNQ 145
OY 284 SDMIATCTTCTTTPPVTTTVSSQSEKSI-----VSGATAPRIVSYOTLN----- 331
DB 146 GKIGIQAPSTELNN-----QKQSDPRAYHNLNLPSSASGTLGSSVPVQGLTKRHPQ 197
OY 332 ----PLAGPVGAKAGVTVLHSV-GPTAATGCTAGTGLLOTSKPLVTSVANTVTVTSIQ 385
DB 198 QMOHPSSSPMYTSG--SFHSFGPNTNAGSTLRLPHLDHSH---MRVHANNQPMGSG 252
OY 386 PEKPVVSGAAVVLSTPAYTFGFTSGAALCLPVKRPVSCWCHICKPVYIGTVQIKLAOP 445
DB 253 LGGPOSTTNM-WTMP--KFERPSS-----VNDRSRV---OG 283
OY 446 GFV--LSOPAGIPTGSSSQQLSLFFHVVOQPSGNEKQVTTISHSSTLLIOKCGOKTMP 503
DB 284 GATSHFQWSSSLPLMSAQGGSSVSHVQESYDOSFEKKNMAASMTSNEDEKSSR-MVL 342
OY 504 NTIIPTSQFPASILKQIT-----LPGNK-----ILS 530
DB 343 ST--PNNMAPASSVSPMTQLDASTMNSRGLGTSGOGANABAPKPKPVSGQKKPLET 400
OY 531 LQASTQKNRIKENTSCPRDE--DDINDVYSMAGVNLNEENACILIA-----TNSLVG 582
DB 401 LGSSPPPSK-KQKVAQSMQDSIDQLNDVYAVSGVNLREEEQLFGAKEDGVRSEASR 459
OY 583 TLIOGCKDEPFLFIALGKRIIDICKKHIDITELNSDVAVLISQAQERLRGLLEKLTALA 642

```

```

DB 460 RVHHEERELLONKPLQKRIAEIKAKGIQISNDYERCLSCVEERNGLSHIITRIS 519
OY 643 QHRMTTYKASENYIILSDPRSOLKPLEKLDLEKQRKDEERMLKAASRSNKEDPQ 702
DB 520 KORVDAEKSHRFTITSIRLQINENMQVKYKEEMKQKQAEKLT---KKPSEKNKEDDK 576
OY 703 LRLKQAKELQQLLELOIQHROANLTATAIG-----PKRRLPLESGIEGLK 749
DB 577 MRTTA-----ANVAAARAAGGDDAFELKQWLMAEARQKSVSEAGKDNQ 619
OY 750 DNLLASGTSLSL-----TATKOLHR-----PRTLR-ICLRDLIFCDEDEREK 790
DB 620 KTTSGGKNSKRODQGRFRSGTGRVGNQSSLIQPKVVTITVQVAVALEREPQNS 679
OY 791 YSRALY 796
DB 680 KSTLMY 685

RESULT 9
OQFMHO
AC Q29071 PRELIMINARY; PRT; 528 AA.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Gastric mucin (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=95275264; PubMed=775593;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lakont J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin.";
RL Biochem. J. 308:89-96(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE-GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA Lakont J.T.;
RT "Pig gastric mucin: isolation and characterization of a cDNA clone
RT with a novel tandem repeat.";
RL Gastroenterology 106:200-200(1994).
DR EMBL; U10281; AAC48526.1; -.
FT NON-TER 1
FT NON-TER 528
SQ SEQUENCE 528 AA; 49907 MW; 0BF0F6879203B2EA CRC64;

Query Match 5.5%; Score 233.5; DB 6; Length 528;
Best Local Similarity 22.3%; Pred. No. 2.1e-05;
Matches 126; Conservative 93; Mismatches 196; Indels 149; Gaps 26;

OY 9 PVSAPKVSQGRPLPAPQIVAAKAPNTTIIQFPAHLQPLPGTVLTKNSGFLMVSPOOT 68
DB 1 PLSVOPSSSSSS--PFTSTTSVQSSSSSVPIPTTSVOP-----SSSGS---APTTS 48
OY 69 VTRAETTSNITSRPAPVAPNPOTVKICTVQPNSSQILIKKAVAPVVKLAQIGTVTVTPK 128
DB 49 ATSVQTS--SSSPIS--TISVQTSSSS-----VPTTSTTSV-Q 85
OY 129 PSSVQAVAPVTSVVT-----VTPGKPLNTVTTLKPSLGSASTPSNEPNLKAENSAVOI 183
DB 86 PSS--SSSAPPTTRATNSVOSSSSSAPISSTTSVQSSSGSVTTTATSVQSSSSSSA--- 140
OY 184 NLSPTMLENVKCKNFMLMLIKLACSGSQSPENGVNKKLVQOLIDAEETFRKLTIV 243

```


Db 141 -----PT-----TSATSVQSSSSSPPISTV-----SVQSSSSSAPTTSATSV 180
 QY 244 ELKSSPQPHLVPFLKKSVALROLPLPNSQSFIOQCVOQTSSDMVATCTT---TTSVP 300
 Db 181 QPSSSSSPPI-----SSTVS-----VOTSSSSVPTTSTTSVQSSSSSV 220
 QY 301 VTTVSSSSQSEKSIIVSGATAPRTVSQTLNPLAGV-----GAKGVYTLHSGV 350
 Db 221 PTTSATSVSRSSSS---SSTPPTSTVSQSSSSSAPTTSATSVQSSSSSPPISTTSVQ 277
 QY 351 PTATGCTTACGTLGLOTS-----KPLVTSVANTVTVSQLOPEKVPVSGRAVT 397
 Db 278 PSSSSSAPTTSATSVQSSSSSPPISTTSVQSSSSSPPISTTSVQSS---SGSAPT 334
 QY 398 LSLPAVTFGETSGAALCLPYSKPVVSCMDHICKPVIGTPOIKLAOPGVLISQAPAGIPT 457
 Db 335 TSATSVQSSSSS-----PPISTTS-----VQSSSSSSPPTTSTTSVQPS 375
 QY 458 GSSSKOLFSLFHVYQOPSGGNEKQVTTISHSSTLTIOKCGQKTMPTVTIIPTSQFPASI 517
 Db 376 SSGSAPTTSATSV---QPSSS-----SSVPTTSATSVSRSSSSSPPIPT---TTSVQPS- 424
 QY 518 LKQITLPGNKILSLQASPTQKNRI 541
 Db 425 ---SSVPTTSATSVQSSSSSPPI 446

RESULT 10

061191 PRELIMINARY; PRT; 2045 AA.

AC 061191;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Transcription factor C1 (HCF).
 GN HCF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-Liver;
 RT "CDNAS encoding the mouse homolog of the human transcription factor C1 (HCF)."
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: us3925; AAB0163.1; .
 DR MGD: MG1:105942; HCF1.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF00041; fn3; 2.
 DR Pfam: PF01344; Kelch; 5.
 DR SMART: SM0060; FN3; 1.
 SQ SEQUENCE 2045 AA; 210537 MW; B89CC2EDA35B969F CRC64;

Query Match

Best Local Similarity 5.5%; Score 233; DB 11; Length 2045;
 Matches 158; Conservative 85; Mismatches 253; Indels 188; Gaps 36;

QY 7 VAPVSAAPKVSQGRRL---PAPQIVAVKAPRTTITQFPANQLDPGTVLIRKNSGPMMLV 63
 Db 486 VLKVTGQATTTGTPFLVYMRASQ---ACKAPVTYV-SLPASVYM---VVPQSQAGTVIGS 539
 QY 64 SPOOT-----VTRAFETSNI---TSRP---AVPAN-----PQYKICTVP 97
 Db 540 NPQMSGAALAAALAAATOKIPSSAPTAWSVPAGTTIVKTVAVTPGTTTLPATVKAASP 599
 QY 98 -----NSSSQLIKKVAV---TPVKKLAQIGTVVTPPKSSVQSAVAVTSVVTYPGAPL 150
 Db 600 VMSNPATRLKTAAPAAQVGSVSAAMTSTRPITTVKSGIV-TVAQAAQVYTVTVGCVT 658
 QY 151 NYVTTLK-PSSL-GASSTPSNEPML-----KAENSAVAQINLSPTMLENVKCKKFL- 200

Db 659 KTIILVKSPIISVPGSALISMLGKMSVQTKPVQTSATVQASTGPTQIITKGPLPA 718
 QY 201 AMLIKLACSSQSP-----EMGQNKVLEBOLLDAKIAEFTTRKLYELKSSPPH 252
 Db 719 GTILKLVTSADGKPTTITTTTQASGAGTKPTILGI-----SSVSPS 759
 QY 253 LVPLKKSVALROLPLPNSQSFIOQCVOQTSSDMVATCTT---SPVVT 302
 Db 760 TT-----KRGTTTIITKIPMSALITQAGATGVTSISGKISPTIITTKVNTSGTAPAKIIT 816
 QY 303 -----TVSSSQSEKSIIVSGA---TAPRT-----VSVQTLNPLAGPVGAK- 340
 Db 817 AVPKATGCGQGVQYVLKAGAPGPGTILRTVPMGVRVLTPTVSAVKRAVATTLVYKG 876
 QY 341 -AGVTLHSGPRTATGTTGTLGLOTSKPLVSA--NVTVVSQLOPEKVPVSGRAVT 397
 Db 877 TTGVTTLGTVTGTVST--SLAGAGHSTASLAPITTLTGITATLSQ---VINPAIT 930
 QY 398 LSLPAVTFGETSGAALCLPYSKPVVSCMDHICKPVIGTPOIKL-AOPGVLISQPA-GI 455
 Db 931 VSAQOTITLTAAGGLTPTITMQPV-----SQPTQVTLITAPSGVEAPVADL 977
 QY 456 PTGSSSKOLFSLFHVYQOPSGGNEKQVTTISHSSTLTIOKCGQKTMPTVTIIPTSQFP 515
 Db 978 PVS-----ILASP-----TTEQPTATVITLADSGQGDVQGVTVLVCNPPC 1018
 QY 516 SI-----LKQITLPGNKILSLQASPTQKNRIENYTSCTREDDDINDYTMAGVNLNE 568
 Db 1019 ETHETGTTNTAVTTVAN--LGHPQPTQVQF-----CDROETASLVTSAVG---Q 1066
 QY 569 ENACILATNSLVGLIOSKDEP 592
 Db 1067 QN-----GNVAVCSNPP 1079

RESULT 11

08TDH7 PRELIMINARY; PRT; 1029 AA.

AC 08TDH7;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Intestinal membrane mucin MUC17 (Fragment).
 GN MUC17.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21845432; PubMed-11855812;
 RA Gum J.R., Jr., Crawley S.C., Hicks J.W., Szymkowski D.E., Kim Y.S.;
 RT "MUC17, a novel membrane-anchored mucin."
 RL Biochem. Biophys. Res. Commun. 291:466-475(2002).
 DR EMBL: AF430017; AAL89737.1; .
 SQ NON-TER 1
 SE SEQUENCE 1029 AA; 108320 MW; 5E1AD19E4BF948E3 CRC64;

Query Match

Best Local Similarity 5.4%; Score 231.5; DB 4; Length 1029;
 Matches 178; Conservative 113; Mismatches 362; Indels 233; Gaps 35;

QY 4 VTKVAPVSAAPKVSQGRRLPAPQIVAVKAPRTTITQFPANQLDPGTVLIRKNSGPMMLV 63
 Db 137 VDRSTPVTTSQSNSTP--TPPEVITLPMSTPSEVSPTLT-MVSTTSVTSISAGTAST 193
 QY 64 SPOOTVTRAFETSNITSRPAPVAPNPTQYKICTVNSSSQLKKVAVPVKKLAQIGTVV 123
 Db 194 LPVDTSTPVTITSTOVSSSPVTPPEG-TTMTPIVT-FSGSTPLTTMPTVSTTRKVTSSGCTLS 251
 QY 124 TVTPKSSVQSAVAVTSVVTYTPGKPLNTVT-----TLKPSLSGASSTPSNEPNIKAENS 178

```

Db 252 T-----PSVVTSTPVTTSTAISATLSDTMSVSMPEISLTITLIVSTTPVTRPES 307
QY 179 AAVOINLSTPTMLENKKCNFLAMLIKACSGSQSPKMGONAKKVLQDLAKIEAEFT 238
Db 308 T-----PSV-----YTSMSMTTASBEGSSP-----TTLLEG 337
QY 239 RKLVELKSSPQHLVPLFKKSVVALRLLPNQSPISQOCVOOTSSDMVATCTTIVTS 298
Db 338 TTMPSMTTERSTLTLTVL-----ISPISWSPSASTLSTPPGDTSTPLATKAGSFSTP 394
QY 299 PVVTTTSSSQSEKSIIVSGATAPRTVSQVTLNPLAGPVGAKAGVTLHVSVPAAATGCT 358
Db 395 AEVTTIRISITERS-----TPLTLIVST-----TL-----PTSPGAS 429
QY 359 TNGTGLQTSKPLVSVANTVTVTSLOPEKPVVSGAVTLSPAVTFGTSCAICLPV 418
Db 430 IASTPLOTSTETPS-----TDTASTPTIPV-----ATTISVSIVTEGSPGTIIPST 480
QY 419 KPVSPSCMD-----HICKPVI-----GTPVQIKLA----- 443
Db 481 -FVTSSTADVPFATTCGAVSTPVITSTELMPTSSSTSTSTSTKEFTTPAATTAAPLT 539
QY 444 -----QPCPVLSOPAGIPTGSSSKOLFSLPHVVO-----PSCGNEKQVTTISHST 490
Db 540 YVTMSTAPSTPRTSRGCTTSASTLSTATSTPHSTSTVTRPVTPSSSSRPSTITSHITP 599
QY 491 LTIQCGCKTMVNTIIPISOPPPA-SILAKITLPGNKILSL-----QASPTQKNRIKEN 544
Db 600 PTPPPASSTPTTSSASTVNPDEAVTMTTRKPTSTRTSTPTVTTTAVPT-NPTIKSN 658
QY 545 VTS-----CFREDDINDVTSMAVNLNENACIATNSLVTGLTIOCKDEPPLF 595
Db 659 PSTPTVPRTTCFGD-----GCQNTASCKCKGWTGDKCOC---PMLT 700
QY 596 IGAQKRI--LDIGKKNIT-----ELNSDAVNLIQATQERLNGLEKTLA 640
Db 701 YELCEEVVSSIDIGPERISAMOMELTVTVSVKFTBELKNISSQOEFQKFTQOMNI 760
QY 641 IAOHKRTTYKASENTILCDTRSOL-KFLEKIDOLEKOKKDLERML--LKAASRSN 696
Db 761 V-----VSGIPEYVGNITKRLGSSVVEHDVLLRTYETPEYKVLDMNATEVVEKEKT 813
QY 697 KEDPEQLRLKO-----KAKELOLELAO-----IOHRDANL 727
Db 814 KYTTOQIMINDCSDMCMCTNTGTQVQNTTVQYDPEEDCKMAKAYGDFVVEYDQR- 872
QY 728 TATTAAGPRKKRPLBSGLKND-----NLASGTSSLATKQLH 767
Db 873 -----PYCISPCEBGFVSXKNKCNLGKQMSLSGPOCLCTTETH 911

```

RESULT 12

```

ID 090WH2 PRELIMINARY: PRT: 2045 AA.
AC 090WH2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE C1 transcription factor.
GN HCFCL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RA Kristie T.M.;
RT "CDNAS encoding the mouse homolog of the human transcription factor C1
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U08921; AAD09225.1; -.
DR MGD; MGI:105942; Hcfcl.
DR InterPro; IPR003961; FN_III.

```

```

DR InterPro: IPR001798; Kelch.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF01344; Kelch; 5.
DR SMART; SM00060; FN3; 1.
SQ SEQUENCE 2045 AA; 210546 MW; 7AD38FCD78EABA9C CRC64;

Query Match 5.4%; Score 231; DB 11; Length 2045;
Best Local Similarity 23.1%; Pred. No. 0.00016;
Matches 158; Conservative 84; Mismatches 254; Indels 188; Gaps 36;

QY 7 VAPVSPKPVSSGPRL---PAPQIVAKAPNTTIOFANLQJPCGVILKSNSGPMLV 63
Db 486 VLKVTGPQATGCTPPLVTRPASQ--AGKAPVYT-SLPASVRA---VPPQSQAGTVIGS 539
QY 64 SPQOT-----VTRAETSTNI--TSRP---AVPAN-----POTVKICTVP 97
Db 540 NPQMSGMAALAAATAAOKIPSSAPLAMSVPAGTTIVKTVAVTPGTTLLPATVKAASSP 599
QY 98 -----NSSOLIKRVAV---TPYKKLAQIGTVVTPPKSSQSVAVPSTVTVTRGKPL 150
Db 600 VMSNPAETRLKTAQAQVGTSSAANTSTRTIITVHKSGTV-TVAQAQAVVTVVVGCVT 658
QY 151 NPVTTLK-PSL-GASTSPNEPNL-----KAENSAVOINLSPMLENKKCNFL-- 200
Db 659 KITTLVKSPTISVPGSALISNLKAVSVQTKPVQTSATVIGASTGPVQIITKGPLRA 718
QY 201 AMILIKACSGOSP-----EMGQNVKLVQDLDAKIEAEFTRLKLVELKSSPOP 252
Db 719 GTILKLVTSADCKPTIITTTQASGAGTKPTILGI-----SSVSPS 759
QY 253 LVFLKKSVALRQLPNQSFIOCCVOOTSSDMVATCTTIVT-----SPVVT 302
Db 760 TT--KPGTTTITKITPMSAIIITQAGATGVTSBPKISPTIITTVMTSGTGAPAKIIT 816
QY 303 -----TVSSSQSEKSIIVSGA-----TAPRT-----VSQVTLNPLAGPVGAK- 340
Db 817 AVPKIKTNGQCGQVTVYVLKAGARGQGTILRTYPMGCVRLVTPVYSAKPAVTLTVKG 876
QY 341 -AGVTLHVSVPAAATGTTAGTGLQTSKPLVTSVA--NTVTVSLQPEKPVSGTAAT 397
Db 877 TIGVTTILGTVGTGST--SLAGAGASTSASLATPITTLCTIATLSSQ---VINPTALT 930
QY 398 LSLPVTPEBTSGAALCLSPKRVVSEFCDHICKPVIIGPVQKL-NQGGPVLSQA-GI 455
Db 931 VSAAGOTTLTAAGGLTPTPTTMOV-----SQPTQVTLTAASGVAQAQVHDL 977
QY 456 PTGSSSKOLFSLFHVVOQPSGNEKQVTTIHSSTLIOCKGKTMVNTIIPISQPPA 515
Db 978 PVS-----LLASP-----TTEQPTATVTINDSGGDVQPGTVILVCSNPPC 1018
QY 516 ST-----LKOITLPGNKILSLQASPTQKNRIKENVTSQFREDDDINDVTSMAVNLNE 568
Db 1019 ETHEGTNTATTTVVAN--LGCHPQPTQOVF-----CDRQETAASLVTSAVG---Q 1066
QY 569 ENACIATNSLVTGLTIOCKDEP 592
Db 1067 QN-----GNVAVCSNPP 1079

```

RESULT 13

```

ID 076602 PRELIMINARY: PRT: 1275 AA.
AC 076602:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 122.9 kDa protein.
GN H02F09.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11

```


RA Cadieu E., Dreano S., Lelaure V., Mottier S., Gallibert F.:
 RT "Sequencing the distal X chromosome of Drosophila melanogaster."
 RN Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Benos P.:
 RT Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003421; AAF4564.1; -
 DR EMBL; AL031028; CA19845.2; -
 DR FlyBase; FBgn0025390; EG:5667.1.
 DR InterPro; IPR002557; Chitin_bind_Pera.
 DR InterPro; IPR002965; P_rich_extensn.
 DR Pfam; PF01607; CBK_14; 2.
 DR PRINTS; PRO1217; PRICHEXTENS.
 DR SMART; SMO0494; ChIBD2; 2.
 SQ SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 5.3%; Score 228; DB 5; Length 1795;
 Best Local Similarity 18.9%; Pred. No. 0.0002;

Matches 172; Conservative 121; Mismatches 367; Indels 250; Gaps 31;

QY 9 PVSAPPKVSGRPLPAPQIVAVKAPNTTITQFPANQL-----PGTVLIRKNSG 58
 DB 700 PVSSTGKPTTPK-PSTRRTPTTTKVTITQTITPLASSFTSTQPTTTPQPTT 758
 QY 59 PLMLVSPQOTVRAETTSNITSRAVPANPOVYKICTVNSSQLIKVAVPVKKLAQI 118
 DB 759 TLLTVTPKSTTTTTEKITSPPK-PTTTOKTSTABNT-----KVAITQKETTPT 812
 QY 119 GTT-----VTVTVPKPS-----VQSAVPTSVTVRGPPLNT 152
 DB 813 OSTSTTITFRKTTTNNPEPTSTKPTSTTPKSTTPKSTTVAASSTKTTTSSPKTTE 872
 QY 153 VTLKRSLSGASTPSEPNLKAENSAVOINLSPMLENVKCKNFMLIKLACSGSQ 212
 DB 873 KSTENPTNSVKTSAVLTSSQRA-----TSTSTPTKTQTQITTTTPKPTLKMS 922
 QY 213 SEPMDGNVKKLYEQLIDAK--IEAEFTKLYELKSSQPHLPVPLKKSVALROLLP 270
 DB 923 TQEATSTQKVSSTVITTTKATESPLTLSTEEPNTTPK-----LRTTTP 970
 QY 271 SOSFIOQCVQOTSDMVIATCTTVTSPVTTVSSSOSEKSIIVSGA--TPAR--TVS 326
 DB 971 TTS-----VATRTTITTTISSESTTSTOKPKSTTPTSTRTTPKTTVY 1017
 QY 327 VQTLNPLAGPVGAKAGVTLHSVG-----PTAATGGTTA-GTGLLOTSKPL 371
 DB 1018 VSTQNPPT--TTSKSTVITITPNPSPSTQRTTTRPTSTIASSTISGTRIPPTTP 1075
 QY 372 VTSVANTVTTVSLQPEKP-----VSGTAATLSLPAYTFCGTSAAICLPSVKPVSF 424
 DB 1076 QNSTSTDLTTVTRPCPDPDSTSDKNTNACTQELQOVNLE-----LQSPKOEQF 1128
 QY 425 CWDHICKPVIGT-----PVQIKLAQPGVLSOPAGIPITGSSSKOFLSF 468
 DB 1129 TTRTRHTALTSRNLIGCEVDPYMDAPSSAFAESQATKAPMSTLAAHLLOKLE 1188
 QY 469 HVV-----QPSGNGEKQVTT-----ISHSSTLTIOKCG 497
 DB 1189 HTSTTPPREHAPTORPSSQSSSQRSGVTAQMARHNLATSKPFIHSLSLTIOQLA 1248
 QY 498 ---QKTMPTNTII-----PTSOFPASILKQITLPGKILS-----L 531
 DB 1249 STQKSIPTKTLVNTHTNTKREPDESEYDSEYDENEVLDTQPRAMSTTVAAYL 1308
 QY 532 QASPTOKNKIKENVTSCFDEDDINDVTMAGV-----LINEACILA 575
 DB 1309 PAVPSTTTRERQKTSSTSPITKATSTTQPTETTGLEDSSSDYVANDAN----- 1363
 QY 576 TNSLVGLTIQSCDEPFLIGALOKRIIDI-----GKKHDTIELNSDAVNLISATQER 630
 DB 1364 ---DISSGVNLSLARKNKLFLSLKQRLQIERTEAKKPATSTSTTADPKTSSSTSP- 1419

QY 631 LRGLLEKLTALIAOHRMTTVKASENYIICSDTRSQLFLEKL--DQLEKQKDLBERML 688
 DB 1420 -----ASTSESTSPVSTYARSKTSLASHKILGEALSRKS-----L 1456
 QY 689 KAKKSRNKKEDPEQLR---LKQAKELQOELALQIQRDNLNLTATNAGRRKKRPLESGI 745
 DB 1457 TPQSAERYDDDDYMEDEPVGSSDAEKKHGVTLISEKQAAATAKNHITAPPSAQPLQAM 1516
 QY 746 EGLKDNLLAS 755
 DB 1517 L-----NILAT 1522

RESULT 15
 ID 015052 PRELIMINARY; PRT; 1246 AA.
 AC 015052;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE KIAA0344 protein.
 GN KIAA0344.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro."
 RL DNA Res. 4:141-150(1997).
 DR EMBL; AB002342; BAA20802.1; -
 SQ SEQUENCE 1246 AA; 129137 MW; 40CEB1209A245C85 CRC64;

Query Match 5.3%; Score 224.5; DB 4; Length 1246;
 Best Local Similarity 19.4%; Pred. No. 0.0002;
 Matches 184; Conservative 133; Mismatches 337; Indels 295; Gaps 40;

QY 9 PVSAPPKVS-SGRPLP-----APQIVAVKAP-----NTTITQFPANQL 45
 DB 340 PNTAPPNFHSHTGPTFPVPPFLSSIAGVPTTAATAPVATSPNDISTSVIGSEVTV 399
 QY 46 LRPQVILINSNGPLMVS---PQOTVTRAEITSNITSRAVPANQYKICL-----Y 96
 DB 400 TEEGIAGVATSTG--VVTSGGLPIPVSESPVLSVVSITIPA---VVISITTSPLQY 454
 QY 97 PNSSOLIKK-----VAVTPVKKILAOIGTVVTVTPKPSVQSAVAPTSVTVTPCKPLN 151
 DB 455 PFTSTSELVSTALYPSVYATSAAGSGTATPGKRPRAVVSQQAAGS---TTVATATL 511
 QY 152 TVTTLKPSLSGASTSPSEPNLKAENSAVOINLSPMLENVKCKNFL-----AM 202
 DB 512 SVST-----TSPFSTASQLSIQLSSPS--FTLAETVVVASHSLDKTSHSSTTG 561
 QY 203 LITL-ACSGSPSEMDGNVKKLYEQLIDAKIEAEFTKLYELKSSQPHLPVPLKKS 261
 DB 562 ASSLAPSSSSSPGACVS-----SYSQPGCLHPLVP---SV 596
 QY 262 VALROLLPNSQSFITQCVQOTSSDMV-----IATCTTVTSPVTTVSSSOSEKSI 315
 DB 597 IASTPLP-----QAAGTSTPLLPQVPSITPLQVPAVNAVQOTLIHSDPQALLP 649
 QY 316 -----VSGVADPRTVSVQTLNPLAGPVGAKAGVTLHSVGPAAFGTGTAGGLQ 366
 DB 650 NQPTHCPPEVSDTOPKACGIDIDIKTLE-----EKLRSILFSEHSSGAQHASVSL 700
 QY 367 TSKRPLTVSAANTVTYVSLQPEKPVVSGTAVTL--SLPAVTREJSGAAICLPSVAPVS 423

```

Db 701 TSLVIESVTBPGIPTTAVAPSKLLSTTSTCLPPTNLPJGT-----VALP-VTPVYT 751
QY 424 FCMWDHCKPVIGTVPVQIKLA--QPG-----PVLISQAPIGTSSSKOLFSLF-- 468
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 752 -----PGQVSTPVSTTSGVKPGTAPSKPPLKAPVLPVGTLPAGTLPBQLPPPPG 804
QY 469 ---HVVOOPSGGNEKQVTTTISHSSTLIQKCGQKTMVNTIIPTSQPPASILKQITLPG 525
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 805 PSLTOSQOPELDDLAQLRRLTSLPEMITV-----TSAVGPVSMAPTAITEAGTQPQ 855
QY 526 NKILSLQASPT-----QKNRIKENVTSCEPDEDDIN---DVTSMAGVNLNEENAC 572
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 856 KGVSVQKEGVPVLATSSGAGVFKMGRFOVSVADGAQKEGKNKSEDAKSVHFESSTSESV 915
QY 573 ILATNSELVGTLIQSCDEPFLFGALQKRILIDIGKKHDITELNSDAVNLIQATQOERLR 632
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 916 LSSSSPE--STLV---KPEP-----NCITIPGISSDVPESAHKTTASEAK 955
QY 633 GLEKLTALIAOHRMTT-----YKASENTILCSDTRSQ-----LKELEKLDL----- 674
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 956 SDTQOPTKVGKRFQVTTANKVGRFSVSKTEBDKITDTKKEGVPVASEPPEMD-LEOAVLPAYI 1014
QY 675 -EKORKDLEEREML-----LKAARS 695
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1015 PKKEPELSEPSHLNGPSSDEPAFLSRVDGSGSPHQLSSKSLPSQNLQSLSNS 1074
QY 696 -----NKEDEPOLRLKOKAKELQLELAQIOHRDAN-----LKAARS 726
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1075 FNSSYMSNDNESDIEDLELRLRNDKHLKEIODLOSROKHIESLYTKLGKVPYAVI 1134
QY 727 LTATAIGPRKKRPLE-----SGIEGLKDNLIASGSSLATKQOLH 767
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1135 IPPAAPLGRRRRPTKSGSKSSRSSSLGKNKSPQLSCNLSGOSASAVLH 1183

```

Search completed: February 16, 2003, 21:58:41
 Job time : 81.8407 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:40:36 ; Search time 27.5208 Seconds
(without alignments)
1738.213 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773
Sequence: 1 GTLVTRVAPVSAAPKVSQGP.....KAGVTLHSVGPATAGTTP 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1773	100.0	801	19 AAW31494	Human hTAFII105 pr
2	1773	100.0	801	22 ABG09468	Novel human diagno
3	1773	100.0	852	21 AAY57279	Transcription facto
4	491	27.7	737	15 AAR56494	TATA-binding prote
5	491	27.7	737	17 AAW6084	Human TATA-binding
6	491	27.7	737	18 AAW25019	Human TATA-binding
7	491	27.7	1023	23 AAU82954	Human homologue of
8	216	12.2	921	15 AAR56487	TATA-binding prote
9	216	12.2	921	17 AAW06077	Drosophila TATA-bl
10	216	12.2	921	18 AAW25028	TATA-binding prote

11	216	12.2	921	22 ABB61528	Drosophila melanog
12	216	12.2	921	22 ABB66055	Drosophila melanog
13	182	10.3	1795	22 ABB69806	Drosophila melanog
14	176	9.9	2035	15 AAR57141	Host cell factor p
15	175.5	9.9	708	22 AAW79978	Human protein SEQ
16	175.5	9.9	709	22 AAW78994	Human protein SEQ
17	175.5	9.9	709	22 AAB93063	Human protein sequ
18	175.5	9.9	709	22 AAB94334	Human protein sequ
19	175.5	9.9	881	22 AAM93811	Human polypeptide,
20	167.5	9.4	5179	22 AAM24516	C899p predicted am
21	164.5	9.3	752	15 AAR51701	AML1-MTG8 fusion.
22	159.5	9.0	557	22 AAB94078	Human protein sequ
23	159.5	9.0	2781	21 AAY57453	Human transcriptio
24	159.5	9.0	2907	21 AAY57452	Human transcriptio
25	156.5	8.8	1328	22 AAM78519	Human protein SEQ
26	156.5	8.8	1331	22 AAM79503	Human protein SEQ
27	156	8.8	1057	22 ABB65440	Drosophila melanog
28	156	8.8	1322	21 AAB42650	Human ORFX ORF2414
29	155.5	8.8	842	22 ABB66631	Drosophila melanog
30	155.5	8.8	864	22 ABB71319	Drosophila melanog
31	153	8.6	612	23 ABB62110	Human prostate spe
32	153	8.6	1296	23 ABB66702	Human novel polype
33	153	8.6	1296	23 ABB66756	Human novel polype
34	152.5	8.6	386	22 ABB29832	Peptide #2483 enco
35	152.5	8.6	386	22 ABB35008	Peptide #2514 enco
36	152.5	8.6	386	22 ABB20422	Protein #2421 enco
37	152.5	8.6	386	22 AAM58821	Human brain expres
38	152.5	8.6	386	22 AAM68195	Human bone marrow
39	152.5	8.6	386	22 AAM16015	Peptide #2449 enco
40	152.5	8.6	386	22 AAM78516	Peptide #2553 enco
41	152.5	8.6	386	22 AAM03749	Peptide #2431 enco
42	152.5	8.6	2870	21 AAY95559	Caenorhabditis ele
43	152.5	8.6	3178	21 AAY95556	Caenorhabditis ele
44	151	8.5	501	19 AAM82571	Human BRP1 DNA
45	150	8.5	571	22 AAB27242	Human EXMAD-20 SEQ

ALIGNMENTS

RESULT 1
AAW31494
ID AAW31494 standard; Protein; 801 AA.
AAW31494:
28-APR-1998 (first entry)
Human hTAFII105 protein.
TATA-binding protein associated binding factor 105; human; activator;
hTAFII105; transcription factor; TFIID; transcriptional activation;
antibodies; diagnosis; therapy; pharmaceutical industry.
Homo sapiens.
US5710025-A.
20-JAN-1998.
02-OCT-1996; 96US-0725012.
02-OCT-1996; 96US-0725012.
02-OCT-1996; 96US-0725012.
(REGC) UNIV CALIFORNIA.
Dikstein R, Tjian R;
WPI: 1998-109818/10.
N-PSDB: AAV02872.
DNA encoding human tata-binding protein associated factor - for
producing recombinant protein

XX Claim 1; Col 17-22; 12pp; English.
XX
XX This cDNA sequence represents a human tata-binding protein associated
CC factor, htaFII105, isolated from Daudi cell nuclear extracts. Tightly
CC associated subunits (TAF's) are components of the transcription factor
CC TFIID and are thought to mediate transcriptional activation. This encoded
CC protein may be produced recombinantly from transformed host cells or
CC purified from human cells. htaFII105 specific binding agents such as
CC specific antibodies could be used for diagnosis (e.g. genetic
CC hybridisation screens for htaFII105 transcripts), therapy (e.g. gene
CC therapy to modulate htaFII105 gene expression) and in the
CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
CC cell specific activators or other transcriptional regulators).
XX
XX Sequence 801 AA;
SQ
Query Match 100.0%; Score 1773; DB 19; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.8e-142;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKVAPVSAAPPKVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGRL 60
DB 1 GTLVTKVAPVSAAPPKVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGRL 60
QY 61 MLVSPQQTVTRAEITTSNITSRPVPANPQTVCITVPNSSQLIKKVAATPVKKLAQIGT 120
DB 61 MLVSPQQTVTRAEITTSNITSRPVPANPQTVCITVPNSSQLIKKVAATPVKKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSIGASPTSPNPNKAENSAA 180
DB 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSIGASPTSPNPNKAENSAA 180
QY 181 VOINLSPTMLENVKCKNFKLMLIKLACSGSQSPKMGQNVKLVLEQLDAKIEAEFTTRK 240
DB 181 VOINLSPTMLENVKCKNFKLMLIKLACSGSQSPKMGQNVKLVLEQLDAKIEAEFTTRK 240
QY 241 LVELEKSSQPHLVPLPKKSVVALROLPLNSQSFIOQCVOQTSSDVAIATCTTVTTSVP 300
DB 241 LVELEKSSQPHLVPLPKKSVVALROLPLNSQSFIOQCVOQTSSDVAIATCTTVTTSVP 300
QY 301 VTTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGTT 359
DB 301 VTTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGTT 359
RESULT 2
ABG09468
ID ABG09468 standard; Protein: 801 AA.
XX
XX ABG09468;
AC
XX
XX 13-FEB-2002 (first entry)
DE
XX Novel human diagnostic protein #9459.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI

XX
XX WPI, 2001-639362/73.
DR N-PSDB; AAS73655.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID No 39827; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp://ipo.int/pub/published_pcl_sequences.
XX
XX Sequence 801 AA;
SQ
Query Match 100.0%; Score 1773; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.8e-142;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTLVTKVAPVSAAPPKVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGRL 60
DB 1 GTLVTKVAPVSAAPPKVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPTVLIKNSGRL 60
QY 61 MLVSPQQTVTRAEITTSNITSRPVPANPQTVCITVPNSSQLIKKVAATPVKKLAQIGT 120
DB 61 MLVSPQQTVTRAEITTSNITSRPVPANPQTVCITVPNSSQLIKKVAATPVKKLAQIGT 120
QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSIGASPTSPNPNKAENSAA 180
DB 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTLKPSLSIGASPTSPNPNKAENSAA 180
QY 181 VOINLSPTMLENVKCKNFKLMLIKLACSGSQSPKMGQNVKLVLEQLDAKIEAEFTTRK 240
DB 181 VOINLSPTMLENVKCKNFKLMLIKLACSGSQSPKMGQNVKLVLEQLDAKIEAEFTTRK 240
QY 241 LVELEKSSQPHLVPLPKKSVVALROLPLNSQSFIOQCVOQTSSDVAIATCTTVTTSVP 300
DB 241 LVELEKSSQPHLVPLPKKSVVALROLPLNSQSFIOQCVOQTSSDVAIATCTTVTTSVP 300
QY 301 VTTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGTT 359
DB 301 VTTTVSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGTT 359
RESULT 3
AA57279
ID AA57279 standard; Protein: 852 AA.
XX
XX AA57279;
AC
XX
XX 06-JUN-2000 (first entry)
DT
XX
XX Transcription factor subunit TAFII105 polypeptide.
DE

XX TATA box-binding protein associated factor II 105; TAFII105; cancer;
KM transcription factor; apoptosis; cytosolic; immunosuppressive;
KM antiinflammatory; virucide; antibacterial.
OS Homo sapiens.
XX Key Location/Qualifiers
FH Misc-difference 802 /label= Glx
FT /note= "encoded by TGA"
FT Misc-difference 834 /label= Glx
FT /note= "encoded by TAA"
XX MO200012699-A1.
XX 09-MAR-2000.
XX 25-AUG-1999; 99WO-IL00464.
XX 27-AUG-1998; 98IL-0125971.
XX (YEDA) YEDA RES & DEV CO LTD.
XX Dikstein R, Yamit-hezi A;
XX WPI: 2000-256640/22.
XX N-PSDB; AA290465.
XX Polypeptide encoding TATA box binding protein associated factor II 105
PT useful for treating e.g. cancers and inducing apoptosis has a dominant
PT negative effect on the normal biological activity of the binding
PT protein -
XX Claim 7; Fig 2; 48pp; English.
XX This represents a polypeptide comprising a (modified) fragment (I) of
CC a TATA box-binding protein associated factor II 105 (TAFII105). A
CC pharmaceutical composition comprising (I) or the polynucleotide or an
CC inhibitor or antagonist of (I) is useful for treating cancers and
CC inducing apoptosis in pathological cells. The composition is also useful
CC for treating autoimmune diseases, inflammatory processes and viral or
CC bacterial infections.
XX Sequence 852 AA:
SQ
Query Match 100.0%; Score 1773; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 8.5e-142;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTATKAPAPAPKSSGRLPAPQIVAVKAPNTTIOFPANQOLPPTVLKINSGL 60
DB 1 GTTATKAPAPAPKSSGRLPAPQIVAVKAPNTTIOFPANQOLPPTVLKINSGL 60
QY 61 MLVSPQQTVAETTSNITSRPAPNPQIVKICTVNSSQLIKKVAVPPVKLAQIGT 120
DB 61 MLVSPQQTVAETTSNITSRPAPNPQIVKICTVNSSQLIKKVAVPPVKLAQIGT 120
QY 121 TVVTVKPPSSVQSAVPTSVTVTPGKPLNTVTTLKPSISGASSTPSNEPNLKAENSA 180
DB 121 TVVTVKPPSSVQSAVPTSVTVTPGKPLNTVTTLKPSISGASSTPSNEPNLKAENSA 180
QY 181 VOINLSTMLNKKCNFLAMLIKACSGSPKMGONKVKLYEQLLDATIEAEFTRK 240
DB 181 VOINLSTMLNKKCNFLAMLIKACSGSPKMGONKVKLYEQLLDATIEAEFTRK 240
QY 241 LVELKSSPQPHLVPLFKKSVVALRLPNSQSFIOQCVQOTSSDMYIATCTTIVTTS 300
DB 241 LVELKSSPQPHLVPLFKKSVVALRLPNSQSFIOQCVQOTSSDMYIATCTTIVTTS 300
QY 301 VTTVSSSQSEKSIIVGATAPRTVSVOPLNLAGPVGAKAGVYTLHSGPTAATGTT 359
DB 301 VTTVSSSQSEKSIIVGATAPRTVSVOPLNLAGPVGAKAGVYTLHSGPTAATGTT 359

DB 301 VTTVSSSQSEKSIIVGATAPRTVSVOPLNLAGPVGAKAGVYTLHSGPTAATGTT 359
RESULT 4
ID AAR56494 standard; Protein; 737 AA.
XX AAR56494;
AC AAR56494;
XX 23-MAR-1995 (first entry)
XX TATA-binding protein-associated factor hTAFII130.
DE TATA-binding protein associated factor; hTAFII130; screening;
KM TATA-binding protein associated factor; hTAFII130; screening;
KW diagnostic; therapeutic; gene transcription regulation.
XX Homo sapiens.
XX OS
XX WO9417087-A.
XX 04-AUG-1994.
XX 28-JAN-1994; 94WO-US01114.
XX 28-JAN-1993; 93US-0013412;
XX 30-JUN-1993; 93US-0087119.
XX (REGC) UNIV CALIFORNIA.
XX Comal L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
XX Tjian R, Wang E, Weinzierl ROJ;
XX WPI: 1994-264019/32.
XX N-PSDB; AAQ70731.
XX TATA-binding protein associated protein factors - and
PT corresponding nucleotide sequence and deriv. antibodies, useful
PT in screening, diagnostics and therapeutics
XX Disclosure; Page 142; 180pp; English.
XX The TATA-binding protein associated factor hTAFII130 (including
CC specific antibodies and fusion products) are used in drug screening,
CC diagnostics and therapeutics. They are used in the development of
CC specific biochemical assays for screening compounds that agonise or
CC antagonise selected transcription factors involved in regulating
CC gene expression associated with human pathology.
XX Sequence 737 AA;
SQ
Query Match 27.7%; Score 491; DB 15; Length 737;
Best Local Similarity 38.4%; Pred. No. 6.9e-33;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
QY 13 PPKVSSG-----PRLPAPQIVAVKAPNTTIOFPANQOLPPTVLKINSGL 61
DB 68 PTKTSGRATLTPTVLAIRLPQ-----PQNTPNIO---NFQPLPMVLYRENGLL 118
QY 62 LVSPQQTVAETTSNITSRPAPNPQIVKICTVNSSQLIKKVAVPPVKLAQIGT 115
DB 62 LVSPQQTVAETTSNITSRPAPNPQIVKICTVNSSQLIKKVAVPPVKLAQIGT 115
QY 119 MI-PQALAQOAOAHQOPQTTMAPRPTSPAPVQISTVQAPGTPIAR-QVT- 172
DB 119 MI-PQALAQOAOAHQOPQTTMAPRPTSPAPVQISTVQAPGTPIAR-QVT- 172
QY 116 AQIGTIVTTPKPSVQSAVPTSVTVTPGKPLNTVTTLKPSISGASSTPSNEPNL 168
DB 116 AQIGTIVTTPKPSVQSAVPTSVTVTPGKPLNTVTTLKPSISGASSTPSNEPNL 168
QY 173 ---TTLIKQV---SQAGTVOFSATLQSPGVOPLVLGGAAGTASLGTAVQGTGPQ 225
DB 173 ---TTLIKQV---SQAGTVOFSATLQSPGVOPLVLGGAAGTASLGTAVQGTGPQ 225
QY 169 NE-PNLKAENSAVAOINLSPMLNKKCNFLAMLIKACSGSPKMGONKVKLYEQL 227
DB 169 NE-PNLKAENSAVAOINLSPMLNKKCNFLAMLIKACSGSPKMGONKVKLYEQL 227
QY 226 RTVGATITSSAATE-----TMEVKKCKNFLTITLIKASSGQSTETANVKEIYONL 279
DB 226 RTVGATITSSAATE-----TMEVKKCKNFLTITLIKASSGQSTETANVKEIYONL 279
QY 228 LDKAIEAEFTTRKLYELKSSPQPHLVPLFKKSVVALRLPNSQSFIOQCVQO---TS 283
DB 228 LDKAIEAEFTTRKLYELKSSPQPHLVPLFKKSVVALRLPNSQSFIOQCVQO---TS 283
QY 280 LDKAIEAEFTTRKLYELKSSPQPHLVPLFKKSVVALRLPNSQSFIOQCVQO---TS 339
DB 280 LDKAIEAEFTTRKLYELKSSPQPHLVPLFKKSVVALRLPNSQSFIOQCVQO---TS 339

```

OY 284 SDWVIATCTTWTSPVTTTSSOSEKSIIVSGATAPRTVS 327
DB 340 Q-----ATTALTAVLSSSVQRTAGKTAATVTSALQPPVLSL 376

RESULT 5
AAW06084
ID AAW06084 standard; Protein: 737 AA.
AC AAW06084;
XX
XX 27-JAN-1997 (first entry)
DE Human TATA-binding protein associated factor hTAFII130 protein.
XX
XX Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
XX RNA polymerase II; transcription; messenger RNA; nuclear fraction;
XX lambda-gt11; expression library.
XX
OS Homo sapiens.
XX
XX US5534410-A.
XX
XX 09-JUL-1996.
XX
XX 28-JAN-1993; 93US-0013412.
XX
XX 28-JAN-1994; 94US-0188582.
XX 28-JAN-1993; 93US-0013412.
XX 30-JUN-1993; 93US-0087119.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Comal L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
XX Tjian R, Wang E, Weinzierl ROJ;
XX
XX WPI: 1996-333245/33.
XX N-PSDB; AAT42217.
XX
XX Screen for cpds. that bind human TATA-binding protein associated
XX factor - by testing ability to bind to polypeptide fragments of the
XX factor, useful as (ant)agonists of transcription factors involved in
XX disease.
XX
XX Examples: Column 105-112; 86pp; English.
XX
XX This is the amino acid sequence of the human TATA-binding protein (TBP)
XX associated factor (TAF) designated TAFII130. The protein is a component
XX of the TFIID fraction required for reconstituting RNA polymerase II in
XX vitro transcription activity. The encoded protein has an estimated mol.
XX wt. of 130 kD by SDS-PAGE.
XX The invention relates to purified proteins involved in transcription
XX by RNA polymerase II, the RNA polymerase which transcribes messenger
XX RNA. RNA polymerase II transcription proceeds in vitro upon addition
XX of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
XX to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
XX contain a TBP and other TAFs. Purification of TFIID and separation of
XX its components reveals 7 proteins ranging in size from 30-250 kD.
XX Serum raised against the TFIID fraction allowed cloning of the corresp.
XX genes from lambda-gt11 expression libraries.
XX
XX Sequence 737 AA:
XX
XX Query Match 27.7%; Score 491; DB 17; Length 737;
XX Best Local Similarity 38.4%; Pred. No. 6.9e-33;
XX Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
XX
XX 13 PKVSSG-----PRAPQVANKAKWTITIQPANLQPPGVYLKKSNGPLM 61
XX |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
XX DB 68 PTAATSGIRATLTPVLARLP-----PONPTNIQ--NFOLEPGVAVLSENGQL 118
XX |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
XX OY 62 LVSPQOTVTR-----AERTSNITSRAVAPANDQIVKICTVNSSQLKKVAVPVKKL 115

```

```

DB 119 MT-PQAKLMOQAHAQPOTTMAPRPAPTSAAPVOISTVQAPGPILAR-OVTP----- 172
OY 116 AQIGTWTVTTPKPSVQSAVPTSVTVPGPLMTVT--TLKPSISGASS-----TPS 168
DB 173 -----TTIIKOV-----SQAGTIVQPSATLQRPSPGVQPOLVLOGAAGTASLGATAVQGTCP 225
OY 169 NE-PNLKAENSAVQINLSPTMLENKKCKANFLAMLKIKACSSGSEPMQNVKKLVEOL 227
DB 226 RTVPGATTTSSAATE-----TMENVKCKKNFLSTLIKIKASSGKSTETAAVKEIVQNL 279
OY 228 LDKAIEAEFTRLVYELKSSPPHLPFLKRSVALROLRPSQSFIOOCVQO-----TS 283
DB 280 LDKRTAEEDFTSRILRELNSSPPQVLPFLKSLPALROLTPDSAAFIQDSQ000PPPPPS 339
OY 284 SDWVIATCTTWTSPVTTTSSOSEKSIIVSGATAPRTVS 327
DB 340 Q-----ATTALTAVLSSSVQRTAGKTAATVTSALQPPVLSL 376

RESULT 6
AAW25019
ID AAW25019 standard; Protein: 737 AA.
XX
XX AAW25019;
XX
XX 08-OCT-1997 (first entry)
XX
XX TATA-binding protein associated factor, hTAFII130.
XX
XX TATA-binding protein associated factor; TAF; nuclear protein;
XX RNA polymerase transcription; TATA-binding protein; TBP;
XX Initiation.
XX
XX Homo sapiens.
XX
XX US5637686-A.
XX
XX 10-JUN-1997.
XX
XX 28-JAN-1993; 93US-0013412.
XX
XX 28-JAN-1994; 94US-0188582.
XX 28-JAN-1993; 93US-0013412.
XX 30-JUN-1993; 93US-0087119.
XX 09-MAY-1996; 96US-0646715.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Comal L, Dynlacht BD, Hoey T, Rupert S, Tanese N;
XX Tjian R, Wang E, Weinzierl ROJ;
XX
XX WPI: 1997-319113/29.
XX N-PSDB; AAT79595.
XX
XX Nucleic acids encoding human TATA-binding protein associated factor
XX (TAF) peptide(s) - for production of recombinant peptide(s), used
XX for modulating transcription of TAFs
XX
XX Claim 1; Column 111-116; 86pp; English.
XX
XX AAW25018 represents TATA-binding protein associated factor (TAF)
XX polypeptide, hTAFII100 (mol. weight 100kD). TAF peptides derived
XX from hTAFII30 alpha, hTAFII30 beta, hTAFII40, hTAFII70, hTAFII100,
XX hTAFII130, hTAFII250, hTAFII48 and hTAFII10 and nucleic acids encoding
XX them, are used to modulate transcription, including transcription
XX initiation. TAFs are nuclear proteins involved in RNA polymerase I,
XX II and III transcription. The peptides act by binding to a different
XX TAF, an activator, or TBP (TATA-binding protein) or competitively
XX inhibiting association of a TAF domain with another compound, typically
XX a protein like TBP or another TAF, an activator, or DNA.
XX
XX Sequence 737 AA:

```

[illegible]

Query Match	27.7%	Score 491	DB 23	Length 1023				
Best Local Similarity	38.4%	Pred. No. 1.1e-32						
Matches 132	Conservative 52	Mismatches 96	Indels 64	Gaps 14				
QY	13	PKVSSG-----	PRLPAPQIVAVKAPNTTTIQPPANQLPQGVILIKSNSGLPM 61					
DB	414	PTATSGIRATLTPVLAPRLPQ-----PONPNIQ-----NFQLPQGVILVRSNGQL 464						
QY	62	LVSPQGYTR-----AETSNITSRPANVPAPQVVKICTVNSSQILKKAIVPVKKL 115						
DB	465	ML-POQALAQMOQAHOAQOTMAPRPAPTPVTSAPPVQISTVQAPGPTPIAR-QVTP---- 518						
QY	116	AOIGTIVVTVYTPKPSVSVAVPTGVVTVTPGKPLNTVY--TLKSSSGASS-----TPS 168						
DB	519	-----TTIIKQV---SQAGTTVPASATLQKSPVQQLVLGGAQTSLSGTAFAVGTGTPQ 571						
QY	169	NE-PLUKENSAVOINISPTMLENVKKCKNFLAMLIKLAGSGSPENQGVKKLVEQL 227						
DB	572	RTVPATTTTSSAATE-----TMENVKKCKNFLTSLIKLAGSGKOSTETAAVAKELVQNL 625						
QY	228	IDAKIEAEFTFKKLVELKSSPQPHLPPLKKSVALRQLPNSQSFIOQCQO-----TS 283						
DB	626	LDGKIEADFTSLRELNSSPQPLVPLPKRSLPALRQLPDSAAFIQ00SQ0QPPPTS 685						
QY	284	SDMVATCTTCTVTPSPVYTTTVSSSQSEKSIIVSGATAPRTVSV 327						
DB	686	Q-----ATTALTAIVLSSSVQRTAGKTAATVTSNALQPPVLSTL 722						
RESULT 8								
AAR56487								
ID	AAR56487 standard; Protein: 921 AA.							
XX	AAR56487:							
XX	23-MAR-1995 (first entry)							
DE	TARA-binding protein-associated factor dTARF1110.							
XX	TARA-binding protein associated factor; dTARF1110; screening;							
KW	diagnostic; therapeutic; gene transcription regulation.							
XX	Drosophila.							
OS	WO94117087-A.							
PN	04-AUG-1994.							
XX	28-JAN-1994;							
PF	94WO-US01114.							
XX	28-JAN-1993;							
PR	30-JUN-1993;							
Sequence 1023 AA;								
Query Match 27.7%; Score 491; DB 23; Length 1023;								
Best Local Similarity 38.4%; Pred. No. 1.1e-32;								
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;								

XX (REGC) UNIV CALIFORNIA.
 XX Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX MPI: 1994-264019/32.
 DR N-PSDB: AAO70724.
 XX TATA-binding protein associated protein factors - and
 PT corresponding nucleotide sequence and deriv. antibodies, useful
 PT in screening, diagnostics and therapeutics
 PS Disclosure: Page 56-61; 180pp; English.
 XX The TATA-binding protein associated factor dTAFII110 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.
 XX

SQ Sequence 921 AA:

Query Match 12.2%; Score 216; DB 15; Length 921;

Best Local Similarity 24.9%; Pred. No. 2.3e-09;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 PQIVAAKAPNTTIOFPANIQLPPTGTVLTKSNGSPLML-----VSPQQTVAETTS 76
 DB 133 PQSPSTLTLNTGQTPA-----LVKTDNGFQLRWGTTGPPYTGITTTNTSNN 184
 QY 77 NITSRAVPANPQVAKICVPPNSSQ-----LIKVAATPVKKLAQIGTTVTTVP 127
 DB 185 NITSTNNHPTTQ-IRLQTVPAASMTNTATSNITVNSVASSGVANSSQPHLTQLMNQ 243
 QY 128 KPSVQSVAVPTSVTVTEGKPLNTVTTLKPSLSGASTPSPNEPNLKAENSAVAQNLSP 187
 DB 244 APOQLPITQITLPAQSOQOQVNNVSSAGTATAVSSTA-----ATT 287
 QY 188 TMLENVK-KCKNFLAMLIKACSGOSPENGQVKKVLBOQLDAKTEAEFTKLYVEIK 246
 DB 288 TQOGNTEKCKRFLANLIEL--STREPKPEKKNVRLIQELVANAVEPEFCRLERLIN 345
 QY 247 SSBPPLVPELKKSVVALRQL-----LPNSQSFQ--- 276
 DB 346 ASBPQLIGFLKSLPLRQALYTKELVTEGKPPQHVGLAGLSQQLPKIQAOIRPIG 405
 QY 277 ----OCVOQTSSDMVATCTTTVTTSPTVTVSSQSEKSIIVSGATAPRTVS---VOT 329
 DB 406 PQGTTTIGQTVAMI--TFNALGTPTPTIGHTTISKQPN---IRLPAPRLVNTGIRI 460
 QY 330 LNPAGPVGAKAGVTVLHSGVPTAA 354
 DB 461 QIP-SLOYVGQANIVQIR--GPOHA 482

RESULT 9

ID AAM06077 standard: Protein; 921 AA.

AAW06077;

27-JAN-1997 (first entry)

Drosophila TATA-binding protein associated factor dTAFII110 protein.

KW Drosophila: TATA-binding protein; TBP associated factor; TFIID;

KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;

XX holoenzyme; lambda-gli1; expression library.
 XX Drosophila melanogaster.

PN US5534410-A.
 XX 09-JUL-1996.
 PD 28-JAN-1993; 93US-0013412.
 PF 28-JAN-1994; 94US-01889582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 XX (REGC) UNIV CALIFORNIA.
 XX Comal L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 DR MPI: 1996-333245/33.
 DR N-PSDB: AAT42210.
 XX Screen for cpds. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.
 XX

Examples: Column 27-36; 86pp; English.

This is the amino acid sequence of the Drosophila TATA-binding protein
 (TBP) associated factor (TAF) designated TAFII110. The protein is a
 component of the TFIID fraction required for reconstituting RNA
 polymerase II in vitro transcription activity. The encoded protein
 has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.
 based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.
 The invention relates to purified proteins involved in transcription
 by RNA polymerase II, the RNA polymerase which transcribes messenger
 RNA. RNA polymerase II transcription proceeds in vitro upon addition
 of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
 to RNA polymerase II holoenzyme. Fraction TFIID has been shown to
 contain a TBP and other TAFs. Purification of TFIID and separation of
 its components reveals 7 proteins ranging in size from 30-250 kD.
 Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gli1 expression libraries.
 CC

SQ Sequence 921 AA:

Query Match 12.2%; Score 216; DB 17; Length 921;

Best Local Similarity 24.9%; Pred. No. 2.3e-09;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 PQIVAAKAPNTTIOFPANIQLPPTGTVLTKSNGSPLML-----VSPQQTVAETTS 76
 DB 133 PQSPSTLTLNTGQTPA-----LVKTDNGFQLRWGTTGPPYTGITTTNTSNN 184
 QY 77 NITSRAVPANPQVAKICVPPNSSQ-----LIKVAATPVKKLAQIGTTVTTVP 127
 DB 185 NITSTNNHPTTQ-IRLQTVPAASMTNTATSNITVNSVASSGVANSSQPHLTQLMNQ 243
 QY 128 KPSVQSVAVPTSVTVTEGKPLNTVTTLKPSLSGASTPSPNEPNLKAENSAVAQNLSP 187
 DB 244 APOQLPITQITLPAQSOQOQVNNVSSAGTATAVSSTA-----ATT 287
 QY 247 SSBPPLVPELKKSVVALRQL-----LPNSQSFQ--- 276
 DB 346 ASBPQLIGFLKSLPLRQALYTKELVTEGKPPQHVGLAGLSQQLPKIQAOIRPIG 405
 QY 277 ----OCVOQTSSDMVATCTTTVTTSPTVTVSSQSEKSIIVSGATAPRTVS---VOT 329
 DB 406 PQGTTTIGQTVAMI--TFNALGTPTPTIGHTTISKQPN---IRLPAPRLVNTGIRI 460
 QY 330 LNPAGPVGAKAGVTVLHSGVPTAA 354

Db 461 QIP-SLOVPGQANIVQIR--GPOHA 482

RESULT 10
AAW25028
ID AAW25028 standard; Protein: 921 AA.

XX
AC AAW25028;
XX
DT 08-OCT-1997 (first entry)
XX
DE TATA-binding protein associated factor, dTAFl110.
XX
KM TATA-binding protein associated factor; TAF; nuclear protein;
KM RNA polymerase transcription; TATA-binding protein; TBP;
XX
XX
OS Drosophila sp.
XX
PN US5637686-A.
XX
PD 10-JUN-1997.
XX
PF 28-JAN-1993; 93US-0013412.
XX
PR 28-JAN-1994; 94US-0188582.
PR 28-JAN-1993; 93US-0013412.
PR 30-JUN-1993; 93US-0087119.
PR 09-MAY-1996; 96US-0646715.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Comal L, Dynlact BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
DR WPI; 1997-319113/29.
DR N-PSDB; AAT79604.
XX
XX
XX Nucleic acids encoding human TATA-binding protein associated factor
PT (TAF) peptide(s) - for production of recombinant peptide(s), used
PT for modulating transcription of TAFs
XX
XX
XX Example 1; Column 35-40; 86pp; English.
XX
XX
XX AAW25028 represents TATA-binding protein associated factor (TAF)
CC polypeptide, dTAFl110 (mol. weight 110KD). TAF peptides derived
CC from dTAFl110 alpha, dTAFl110 beta, dTAFl140, dTAFl160, dTAFl180,
CC dTAFl1110, dTAFl150, and dTAFl1250, their human equivalents and
CC nucleic acids encoding them, are used to modulate transcription,
CC including transcription initiation. TAFs are nuclear proteins involved
CC in RNA polymerase I, II and III transcription. The peptides act by
CC binding to a different TAF, an activator, or TBP (TATA-binding protein)
CC or competitively inhibiting association of a TAF domain with another
CC compound, typically a protein like TBP or another TAF, an activator,
CC or DNA.
XX
XX
XX Sequence 921 AA;

Query Match 12.28; Score 216; DB 18; Length 921;
Best Local Similarity 24.98; Pred. No. 2.3e-09;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 PQYAVAKAPNTTITQFANLQLPGTVLILKNSGPIML-----VSPQQTVAERTS 76
DB 133 PQSSITLSTLNTGQTPA-----LLVKTDFOLLRVTTGPPVYOTITINTSINS 184
QY 77 NITSRAVPANPOTVKICVYVNSSQ-----LIRKVAVTPYKKAQIGTVVTVTP 127
DB 185 NTSTSTNHPPTTQ-IRLQTVPAASMTNTATSNIIYNSVASSGVANSSQPHLTLQNAQ 243
QY 128 KPSSVGVAVPTSVYVTPGKPLNTVTTIKRSSIGASSTPENENLKRKENSAAVQINISF 187
DB 244 APQLPQITQITQIPAGOSQOOOVNNSVSSAGCTAAVASSTTA-----ATT 287

QY 188 TMLENVK-KCKNFLAMLIKACSGSPDMGONVAKLVEQLIDAKIEAEFTRKLYELK 246
DB 288 TQGGTKKCKRKFLLANLEL---STREPKVKEKNVTLQELVNAVVEFEEDRERLLN 345
QY 247 SSPQPHLVPLKKSVALRQL-----LPSQSFQ--- 276
DB 346 ASPQCLIGFLKKSILPLLRLQALYFKELVIEGIPPPQVHLGAGSQOLPKIOAIRPIG 405
QY 277 -----QCVQQTSSDMVIACTTIVTSPVYTTVSSQSEKSIYSGATAPRTVS---VQT 329
DB 406 PSQTTTIGQTVRMT--TPNALGTPRPRIGHTTISKQPN--IRLPAPRLVNTGIRT 460
QY 330 LNPAGPYGAKAGVYTLHSGPTAA 354
DB 461 QIP-SLOVPGQANIVQIR--GPOHA 482

RESULT 11
ABB61528
ID ABB61528 standard; Protein: 921 AA.

XX
AC ABB61528;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11376.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX
XX (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
DR N-PSDB; ABL05631.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
XX Disclosure; SEQ ID NO 11376; 21pp + Sequence Listing; English.
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 921 AA;

Query Match 12.28; Score 216; DB 22; Length 921;
Best Local Similarity 24.98; Pred. No. 2.3e-09;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIYAVKAPNTTIOFPAQLPFGTVLIRKNSGRLML-----VSPQOTVTRAETTS 76
 DB 133 POSPSTLTLTNTGQTPA-----LVKTDNGFOLLRGTTGTPPTVTOTITNTSNNS 184
 QY 77 NITSRAVPANPQTVKICITVPNSSQ-----LIKVAVTPVKKLAQIGTVVTVTP 127
 DB 185 NTSTSTNHPTTQ-IRLQTVPAASMTNTATSNIIYNSVASSGYANSSQPHILQIQA 243
 QY 128 KPSVQSVAVPTSVYVTPKPLNTVTTLKPSLSGASSTPSNEPNLKAENSAVAQINLSP 187
 DB 244 APOLPQITQITIPAOOSQOQOVNNSVAGGTATVAVSSTTA-----ATT 287
 QY 188 TMLENVK-KCKNPLAMLIKLAGSGSOSPEMGONVKLVLEQLDAKTEAEETRKLYELK 246
 DB 288 TQGNTEKCKRKLNLIEL--STREPKVEKNVRLIOELVNVANVEPEFCORLERLN 345
 QY 247 SSPQHLVPELKKSVVALROL-----LPNSQSFIO--- 276
 DB 346 ASPQCLIGFLKSLPLRLQALYTKELVIGIKRPPQHVLAGLSQQLPKIQAQIRPIG 405
 QY 277 ----QCVOQTSSDMVATCTTTVTTSPPVYTTVSSOSSEKSTIVSGATAPRTVS---VOT 329
 DB 406 PSQTTTIGQTVRMV--TPNALGTPRTTIGTHTTISKQPPN--IRLPTARLVNTGQIRP 460
 QY 330 LNPAGPVGAKAGVYTLHSVGTAA 354
 DB 461 QIP-SLOVPGQANIVQIR--GPOHA 482

RESULT 12

ABB66055
 ID ABB66055 standard; Protein: 921 AA.

AC ABB66055;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 24957.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL10158.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 24957; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins

CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 921 AA;

Query Match 12.2%; Score 216; DB 22; Length 921;
 Best Local Similarity 24.9%; Pred. No. 2.3e-09;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIYAVKAPNTTIOFPAQLPFGTVLIRKNSGRLML-----VSPQOTVTRAETTS 76
 DB 133 POSPSTLTLTNTGQTPA-----LVKTDNGFOLLRGTTGTPPTVTOTITNTSNNS 184
 QY 77 NITSRAVPANPQTVKICITVPNSSQ-----LIKVAVTPVKKLAQIGTVVTVTP 127
 DB 185 NTSTSTNHPTTQ-IRLQTVPAASMTNTATSNIIYNSVASSGYANSSQPHILQIQA 243
 QY 128 KPSVQSVAVPTSVYVTPKPLNTVTTLKPSLSGASSTPSNEPNLKAENSAVAQINLSP 187
 DB 244 APOLPQITQITIPAOOSQOQOVNNSVAGGTATVAVSSTTA-----ATT 287
 QY 188 TMLENVK-KCKNPLAMLIKLAGSGSOSPEMGONVKLVLEQLDAKTEAEETRKLYELK 246
 DB 288 TQGNTEKCKRKLNLIEL--STREPKVEKNVRLIOELVNVANVEPEFCORLERLN 345
 QY 247 SSPQHLVPELKKSVVALROL-----LPNSQSFIO--- 276
 DB 346 ASPQCLIGFLKSLPLRLQALYTKELVIGIKRPPQHVLAGLSQQLPKIQAQIRPIG 405
 QY 277 ----QCVOQTSSDMVATCTTTVTTSPPVYTTVSSOSSEKSTIVSGATAPRTVS---VOT 329
 DB 406 PSQTTTIGQTVRMV--TPNALGTPRTTIGTHTTISKQPPN--IRLPTARLVNTGQIRP 460
 QY 330 LNPAGPVGAKAGVYTLHSVGTAA 354
 DB 461 QIP-SLOVPGQANIVQIR--GPOHA 482

RESULT 13

ABB69806
 ID ABB69806 standard; Protein: 1795 AA.

AC ABB69806;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 36210.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL13909.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
XX
PS Disclosure; SEQ ID NO 36210; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 1795 AA;

Query Match 10.3%; Score 182; DB 22; Length 1795;
Best Local Similarity 22.1%; Pred. No. 4.7e-06;
Matches 86; Conservative 52; Mismatches 165; Indels 86; Gaps 13;

QY 9 PVSAAPKVSQPRLPAPQIVAVKAPNTTIOFPANLQI-----PGGVLIKNSG 58
DB 700 PPSSTGKPTTPK-PSRTTPTTKVTTTQITTTPLRSSSTETTSTQPTTTPQPTT 758
QY 59 PLMLVSPQOVYRAETTSNITSRAVPANPQTVKICTVPSNSSOLIKKVAVTPVKILAOI 118
DB 759 TLLVTPKSTSTTTTTEKPTSSPK-PTTTOQKTSFAPNTT-----KVALTTQETPT 812
QY 119 GTT-----VTTVPKPS-----VQSAVPTSVTVTPCKPLNT 152
DB 813 QSTSTTTFRTKTTNNPEPTSTKPTISTPKPSTTPPKSTVASSTKTTISSPKPTTE 872
QY 153 VTTLPKPSISASTPSNPNKKAENSAVOINLSPTMLENVKCNFLAMLIKLAGSGSQ 212
DB 873 KSTENPTTNSKTSALNISTORA-----ISTSEPTKTQNTTTTPKPTTKTS 922
QY 213 SPEGQNVKLVLEQLDAK--TEAEFTRLKLVLEKSSPOHVLPELKKSVVALKQLLPN 270
DB 923 TQEAFTTSQKSVTVITTKKATESPLTTLSTEBNTPKP-----LRTTPT 970
QY 271 SOSITQOCVOQTSSDMVATCTTVTTSPPVTTVSSQSEKSIIVGCA--TAPR--TVS 326
DB 971 TTS-----VATRTTITTTTISESTETTSQKPKSTPSTRTTPKVTVI 1017
QY 327 VQTLNPLAGPVGAKAGVTVLHSGVPTAT 355
DB 1018 VSTQNPPT--TTSKTSVTTTTPNPSPT 1044

RESULT 14
AAR57141
ID AAR57141 standard; Protein; 2035 AA.
XX
AC AAR57141;
XX
DT 19-MAR-1995 (first entry)
XX
DE Host cell factor protein.
XX
KW Herpes simplex virus; herpes virus; VP16; immediate early gene;
KW host cell factor; virus infection therapy; cellular protein;
KW transcription.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Active-site 21..31
FT Active-site /note= "peptide R60"
FT Active-site 168..186
FT Active-site /note= "peptide R37"
FT Active-site 333..340

FT Active-site /note= "peptide R52"
FT 426..449
FT /note= "peptide 362"
FT 511..526
FT /note= "peptide 329"
FT 578..594
FT /note= "peptide 223 first sequence"
FT 594..611
FT /note= "peptide R26 1st peptide"
FT 611..623
FT /note= "peptide 223 2nd sequence"
FT 723..731
FT /note= "peptide 318"
FT 802..813
FT /note= "peptide 299"
FT 813..820
FT /note= "peptide 268"
FT 836..847
FT /note= "peptide R26 2nd sequence"
FT 1010..1031
FT /note= "THE TNT repeat 1"
FT 1072..1093
FT /note= "THE TNT repeat 2"
FT 1101..1126
FT /note= "THE TNT repeat 3"
FT 1158..1183
FT /note= "THE TNT repeat 4"
FT 1286..1311
FT /note= "THE TNT repeat 5"
FT 1314..1339
FT /note= "THE TNT repeat 6"
FT 1349..1374
FT /note= "THE TNT repeat 7"
FT 1414..1439
FT /note= "THE TNT repeat 8"
FT 1774..1781
FT /note= "peptide 293 2nd sequence"
FT 1808..1819
FT /note= "peptide 115"
FT 1819..1840
FT /note= "peptide 261 1st sequence"
FT 1853..1863
FT /note= "peptide 240"
FT 1901..1919
FT /note= "peptide R32"
FT 1919..1930
FT /note= "peptide 261 2nd sequence"
XX
XX W09413315-A.
XX
XX 23-JUN-1994.
XX
XX 03-DEC-1993; 93WO-US11721.
XX
XX 04-DEC-1992; 92US-0989842.
XX 12-APR-1993; 93US-0046585.
XX
XX (COLD-) COLD SPRING HARBOR LAB.
XX (TULA-) TULARIK INC.
XX
XX Herr W, Lamarco K, Wilson A;
XX
XX WPI; 1994-234207/28.
XX N-PSDB; AAQ69229.
XX
XX New Host Cell Factor polypeptide(s) and nucleic acid - are used
XX to develop agents for diagnosis or treatment of disease
XX associated with expression of a HCF-modulated gene e.g. viral
XX infections
XX
XX Disclosure; Page 39; 71pp; English.
XX
XX HCF is required for the transcription of a number of
CC

CC viral genes, such as the immediate early herpes simplex virus-1
CC genes. Epitopes of the encoded protein can be used in
CC defining functional domains of HCF. Identifying compounds that
CC associate with HCF or designing compounds capable of modifying HCF
CC transcription. Such agents can be used to treat viral infections.

SO Sequence 2035 AA:

Query Match 9.9%; Score 176; DB 15; Length 2035;
Best Local Similarity 24.2%; Pred. No. 1.8e-05;
Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

QY 10 VSAPRVSSGPRLPAPQIYAVKAPNTTIOFPANLQPEPTVLKISGCLMLVSPQOTV 69
DB 537 IGSSPQMSGMAALAAAAATOKIPSSA---PTLVSPACTTIVKT---MAVTPQT 588
QY 70 TRAEITSNITSRPAVPANPOT-----VKICTVPSNSSO---LIKKVAVTPYK 115
DB 589 LPA--TVKVAASPRVWWSNPATRLKTAAGVGTSSSATPTSTRPIITVHKSGTVVAQ 646
QY 116 AQTGTTV-----TTVPKP-----SSQSVAVPTS----- 140
DB 647 AGVTTTVGGVTKITLVKSPISVPGSALISMLGKMSVYQTKPVQTSVAVTQASTGPV 706
QY 141 -----VTVTPGKPLNVTY-----LKPSGLCASST-----PSNE 170
DB 707 TQIIQKGPLPACTILKLVTSADGKPTTITTTQASGAGTKPILGISVSPSTTKRG 766
QY 171 PNKAENSAVAQINLSPMLNKKCKNFLAML--IKLACSGSQSPKMGQVKKVLEQ 229
DB 767 TIKTIPMSLITQAGATGVTSSPGIKSPITTTTKVMTSGTAP-----AKIITAV-- 818
QY 230 AKIEAEETFRKL-YVEIKSSP-QP---HLVPLKRSVALROLPLNSQSFIOCVQOTS 283
DB 819 PKATGHHGQGVQVVLKGAFCGPGLIRTP-----MGVRLVTPVYSAVKPAV---- 869
QY 284 SDNVIAFCTTPTVTSPVTTTSSSOS-----EKSIIVSGATFARTVSVQTLNP 332
DB 870 -TTLVVKGTGVTTLGTVGTSTSLAGAGHSTASLATPITTLGITA--TLLSOVINP 926
QY 333 LAGPVGAKAGVTLHSGVPAATAGTT 359
DB 927 TA--ITVSAAGTTL-----TAAGGLTT 946

RESULT 15
AAM79978
ID AAM79978 standard; Protein; 708 AA.
XX
AC AAM79978;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3624.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001MO-US04098.
XX
PR 03-FEB-2000; 2000US-0496814.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
DR WPI: 2001-476283/51.
DR N-PSDB; AAK53111.

CC Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX

PS Claim 20; Page 401; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX

SO Sequence 708 AA:

Query Match 9.9%; Score 175.5; DB 22; Length 708;
Best Local Similarity 25.4%; Pred. No. 4.4e-06;
Matches 106; Conservative 53; Mismatches 166; Indels 93; Gaps 20;

QY 3 LTVKVAVPSAPR-----VSSGPRLPAPQIYAVKAPNTTIT---QPPAN----- 43
DB 239 LVTPRAVAVSSQPKLQRPVTSGLTATSV--LAPNATVAVATVQVSGNRPQTSLQPL 295
QY 44 -----LQLPFGTVLKSNGSPMLVSPQOTVRAETTSNITSRPAVPAN 87
DB 296 PVILHNPVAVSSQPKLQRPVTSGLTATSV--LAPNATVAVATVQVSGNRPQTSLQPL 354
QY 88 PQTVKICTVNSSQLKKKAAVTPVKLAQIGTTV--VTYPRKSSQSVAVPTSVYTP 146
DB 355 P--TRNNVPSVPSIQR--NPTASAPALGTTLLAQAAPVTAHSIVQ--VTRTSLPTVGP 408
QY 147 GKPLNTVTLKP-----SSLGASSTPSNEPLKANSAAVQINLSPMLEVVK 195
DB 409 SGLYSPETNGPIQOMKIPISAFSTSSAAEQNSNTPTREINQTKTIDASVSKRAASTSQ 468
QY 196 CKNFLAMLKLCSSGOS-----PMG--QNVKKVLEQLDAKIEAEETFRKLYV 243
DB 469 CGR-----ATGSDSSGVIDLTMDEESGASQDKIKNHIPVSTWSSSQVSRPLQ 519
QY 244 ELKSSP-QPHLVPLFKRSVALROLPLNSQSFIOCVQOTSSDMVATCTCTTYTSP--- 299
DB 520 IQPAPPLQPSGVPSTGSGQTTI--HLPTAPY---TVNVTHRPVTOVTRLPVPRAVANH 574
QY 300 -VYTTVSSSSQSEKSI--IVSGATAPRTVSVQTLNPLAGVGAAGV---VTLHSGVP 351
DB 575 QVYVYTTLPAPPAQAPLNGTVMQAPAVAGVAPQNSVTVRVQTTTYVNNGLIGSGCP 632

Search completed: February 16, 2003, 21:55:06
Job time : 34.5208 secs

Query Match 27.7%; Score 491; DB 1; Length 737;
 Best Local Similarity 38.4%; Pred. No. 7.9e-37;
 Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PPKVSSG-----PRIPAOIVAVKAPNNTTIOFPANLQIPCTVLKNSGFLM 61
 Db 68 PTAATSGIRATLPTVLAPRLPQ-----PONPTNID---NFOAPPKMLVRSNGQL 118
 QY 62 LVSPQOQVTR-----AETSNITSRAPVAPNPOVTKICTVFNSSSOLIKKAVVPPVKL 115
 Db 119 MI-PQOALAOQAOAHAPOTTMAPRPPTTSAPPVQSTVOAECTPIIAR-QVTP----- 172
 QY 116 AQIGTAVTVTPKPSVQSAVAVPTSVTPGKPLNTVT--TLKPSLIGASS-----TPS 168
 Db 173 ----TTIKQY---SQAQTVQPSATLQRPSPGVQPOLVIGMAQASIGTATAVQGTGPQ 225
 QY 169 NE-FRLKRNENAAVOINLSPTMLENVKCKNFAMLIKACSSGSPSEMGONVKKLVQL 227
 Db 226 RTVPGATTTSSAATE-----TMENVKCKNFSLTLIKASSGKSTETANVKKELVQNL 279
 QY 228 LDKIEAEFTFKLYVELKSSPOPVLVPELKSVYALROLPLNSQSFIOCCVQO-----TS 283
 Db 280 LDGITEADEFTSKLYRELNSSPQPLVPELKRSPLALQPLPDSAAFTQOQSOQOPPPPTS 339
 QY 284 SDWVIATCTTPTVTSPVTTVTSSQSEKSIIVSGATAPRTVS 327
 Db 340 Q-----ATTALAVVLSSSVQRTAGKTATVTSALQPPVLSL 376

RESULT 4
 US-08-188-582-2
 ; Sequence 2, Application US/08188582
 ; Patent No. 5534410

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
 APPLICANT: Comai, Lucio
 APPLICANT: Dynlacht, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoko
 APPLICANT: Wang, Edith
 APPLICANT: Weinzierl, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/188,582
 FILING DATE: 28-JAN-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-57650-2/AUT/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 921 amino acids

TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-188-582-2

Query Match 12.2%; Score 216; DB 1; Length 921;
 Best Local Similarity 24.9%; Pred. No. 2.6e-11;
 Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIVAVKAPNNTTIOFPANLQIPCTVLKNSGFLM-----VSPQOVTFAETTS 76
 Db 133 PPSPTSLTSLNTGQTPA-----LVKTDNGFQLRVGTTGPPVPTTITTSNNS 184
 QY 77 NITSRAVAPNPOVTKICTVFNSSSQ-----LIKKVAVTPYKKAQIGTAVTVTP 127
 Db 165 NITSTNNPTTQ-IRLQTVPAASMTNTTATSNIIIVASVSSGVANSQPHPLQLMAQ 243
 QY 128 KPSVQSAVAVPTSVTPGKPLNTVTTLKDSLIGASSSTPSNEPNLKAENSAVOINLS 187
 Db 244 APQLPQIQIOTIPAQSQOQOQVNNVSSAGATAVASTTA-----ATT 287
 QY 188 TMLENVK-KCKNFAMLIKACSSGSPSEMGONVKKLVQLLDKIEAEFTFKLYVELK 246
 Db 288 TQGNFKKCKRFLANLIEL--STREPKVEKNVFTLQELVNAVPEEFCDRLERLN 345
 QY 247 SSPQHLVPELKSVYALROL-----LPNSQSFIO--- 276
 Db 346 ASPQCLIGFLKSLPLRQALYTKELYEGIKPPQHVGLAGLSQQLPKQAQIRPIG 405
 QY 277 ---OCVOQTSDDWVIATCTTPTVTSPVTTVTSSQSEKSIIVSGATAPRTVS---VOT 329
 Db 406 PSQTTTIGQTVGRMI--TPNALGTPRPITGHTTISKQPPN--IRLPLRVLNTGIR 460
 QY 330 LNPAGVYKAGAVTTLHSGPTAA 354
 Db 461 QTP-SLQVPGQANVQIR--GPOHA 482

RESULT 5
 US-08-646-715-2
 ; Sequence 2, Application US/08646715
 ; Patent No. 5637686

GENERAL INFORMATION:

APPLICANT: Tjian, Robert
 APPLICANT: Comai, Lucio
 APPLICANT: Dynlacht, Brian D.
 APPLICANT: Hoey, Timothy
 APPLICANT: Ruppert, Siegfried
 APPLICANT: Tanese, Naoko
 APPLICANT: Wang, Edith
 APPLICANT: Weinzierl, Robert O.J.
 TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
 NUMBER OF SEQUENCES: 36
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/646,715
 FILING DATE: 09-MAY-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/188,582
 FILING DATE: 28-JAN-1994

ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-715-2

Query Match 12.2%; Score 216; DB 1; Length 921;
Best Local Similarity 24.9%; Pred. No. 2,6e-11;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 PQIVAVKAPNTTITIOFPANLQLPPTGVILIKNSGPIML-----VSPQOVTTRAETTS 76
DB 133 POSPSTILSLNTNGQTPA-----LLVKTONGFOLLRVGTTGPTVTQTITNTSNN 184
QY 77 NITSRVAVNPQTVKICTYPNSSQ-----LIKAVTPYKKAIGITVTVP 127
DB 185 NNTSTNHPPTTQ-IRLQVPAASMTNTATSNITVNSVASSGYANSSOPHILQOLNAQ 243
QY 128 KPSSVAVAPTSVTVTPPKPLNTVTLKPSISGASTPSNEPNLKAENSAVQINLSP 187
DB 244 APQLPQITQITIPAOSSQOQYNNVSSAGTATVASTTA-----ATT 287
QY 188 TMLNVK-KCNFLAMLIKACSGSPENGQVKKLVEQLDAKIEAEFTKLYELK 246
DB 288 TQGNMKEKCKFLANLIEL-STREPKVEKKNVRLIQLVNAVNEPEECRLLELLN 345
QY 247 SSPQPHVPLKKSVALRDL-----LNSQSFQ--- 276
DB 346 ASPOPLIGLKSLPLKRALYKELVIGCIRPPQHVIGLAGLSQQLKIDQIKRPIG 405
QY 277 ---QCVOQTSDMVIATCTTVTTPVVTTVSSQSEKSIIVSGATAPRTVS---VOT 329
DB 406 PSQTTTIGQTVBMI--TPNALGTPTPTIHTTISKQPN---IRLPTARLVNTGIRI 460
QY 330 LNPAGVCAKAGVTLHSVPTAA 354
DB 461 QLP-SLQVPGQANIVQIR--GPOHA 482

RESULT 6
US-08-046-585-5
Sequence 5, Application US/08046585
Patent No. 5453362
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
TITLE OF INVENTION: HOST CELL FACTOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/046, 585
FILING DATE: 12-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-046-585-5

Query Match 9.9%; Score 176; DB 1; Length 2035;
Best Local Similarity 24.2%; Pred. No. 4,3e-07;
Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

QY 10 VSAPKVSQPRPLPAPQIVAVKAPNTTITIOFPANLQLPPTGVILIKNSGPIMLVSPQOIV 69
DB 537 IGSSPQMSGMAALAAAAATOKIPSSA---PYLSVPACTTIVKT-----MAVTPTT 588
QY 70 TRATTSNITSRPVAVNPOT-----VKICVPPNSSQ---LIKAVTPYKKL 115
DB 589 LPA--TVKAVSPVMSNPATRLKLTAAAOVGISVSSATITSPITTVKSGTIVYAOQ 646
QY 116 AOIGTIVV-----TVPRP-----SSVQAVNPTS----- 140
DB 647 AOVTVVGVGTITILVKSPIVPGSALISNLGVSVQKPVQTSVAVTGASTGPV 706
QY 141 -----VIVTPKPLNTVTT-----LKPSSIGAST-----PSNE 170
DB 707 TQIIQKPLPACTILKLVISADCKPTTITTTQASGAGTKPPTLIGISSVPSPTKGT 766
QY 171 PNKAENSAVQINLSPTMLNVKCKCNFLAML-IKACSGSOPENGQVKKLVEQLD 229
DB 767 TIKTITMSALITQAGATGTSRGIKSPITITTKYMTSGTAP-----AIIIV-- 818
QY 230 AKIEAEFTKRL-VYELKSSP-QP---HLVPLKKSVALROLPLNSQSFIOQVOQTS 283
DB 819 PKIATGCGQGVQVYLKAPGCGPTILRTVP-----MGVRLVTPVTVSAVPAV---- 869
QY 284 SDMVIATCTTVTTPVTTVSSQS-----EKSIIVSGATAPRTVSQTLNP 332
DB 870 -TTLVKGITGVVTLGTVGTIVSTSLAGAGHSTASLSLATPITTLGTVIA--TLLSOVINP 926
QY 333 LAGPVCAKAGVTLHSVPTAATGTT 359
DB 927 TA--ITVSAQITL-----TAAGGLIT 946

RESULT 7
US-08-393-703-5
Sequence 5, Application US/08393703
Patent No. 5585239
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
TITLE OF INVENTION: HOST CELL FACTOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA

```

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,703
FILING DATE: 24-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-2/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-393-703-5

```

```

Query Match          9.9%; Score 176; DB 1; Length 2035;
Best Local Similarity 24.2%; Pred. No. 4.3e-07;
Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

```

```

QY 10 VSAPPKVSSEGRPLPAPQVAVKAPNTTIOFPANQLPFGVILKNSGPMILVSPQOTV 69
DB 537 IGSSPQKSGMALALAAAATOKIPSSA---PYLVSPAGTTIVKT-----MAVPGTGT 588
QY 70 TRAEFTSNITSRPAVPANPQT-----VKICTVPNSSQ---LIRKVAATPVKRL 115
DB 589 LPA--TYVASSPVVWNSPATRMLKTAAGVGTSSATNTSTRPIITVHKSGVTVAOQ 646
QY 116 AQIGTTVV-----TVPKP-----SSQVAVVPTS-----140
DB 647 AQVTVVGVGVTKTITLVKSPISVPGSALISNLKVMNVQTKPVQTSAVTGAQTGPV 706
QY 141 -----VVTYTPGKPLNTVTT-----LKPSSIGAST-----PSNE 170
DB 707 TQIIQTKGPLPAGTILKLVTSADGKPTIITTTQASGAGTPTIIGTSSVSPSTTKPGTT 766
QY 171 PNLKAENSAVQINLSPTMLENVKCKNFKLAML-IKLACSGSOSPQEMQNVKRLVEQLLD 229
DB 767 TIITIPMSALITQAGATGVTSPEIKSPIITITTKVMTSGTGA-----AKIITAV-- 818
QY 230 AKIAEEFTRL-YVELKSSP-QP-----HLVPFLKSVVALROLNPSQSFIOOCVOQTS 283
DB 819 PKIATGHQOGCVTVLKGAGQPGTILRTVP-----MGVRLVTPVVSVAVKPVP----- 869
QY 284 SDMYIATCTTIVTSPVTTVTVSSQS-----EKSIIVSGATAPRTVSQVLNP 332
DB 870 -TTLVVKGTGTGVTGLGVTGTGVSISLAGAGHSTASLAPITTLGLTIA--TLSSQVNP 926
QY 333 LAGPVGAKAGVTLHSVGPATAGGTT 359
DB 927 TA--ITVSAQOTL-----TAAAGLTT 946

```

```

RESULT 8
PCT-US93-11721-5
Sequence 5, Application PC/TUS9311721
GENERAL INFORMATION:
APPLICANT: Lamarco, Kelly
APPLICANT: Wilson, Angus
APPLICANT: Herr, Winship
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:

```

```

TITLE OF INVENTION: HOST CELL FACTOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11721
FILING DATE: 03-DEC-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: EP-57503-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-11721-5

```

```

Query Match          9.9%; Score 176; DB 5; Length 2035;
Best Local Similarity 24.2%; Pred. No. 4.3e-07;
Matches 108; Conservative 57; Mismatches 148; Indels 134; Gaps 22;

```

```

QY 10 VSAPPKVSSEGRPLPAPQVAVKAPNTTIOFPANQLPFGVILKNSGPMILVSPQOTV 69
DB 537 IGSSPQKSGMALALAAAATOKIPSSA---PYLVSPAGTTIVKT-----MAVPGTGT 588
QY 70 TRAEFTSNITSRPAVPANPQT-----VKICTVPNSSQ---LIRKVAATPVKRL 115
DB 589 LPA--TYVASSPVVWNSPATRMLKTAAGVGTSSATNTSTRPIITVHKSGVTVAOQ 646
QY 116 AQIGTTVV-----TVPKP-----SSQVAVVPTS-----140
DB 647 AQVTVVGVGVTKTITLVKSPISVPGSALISNLKVMNVQTKPVQTSAVTGAQTGPV 706
QY 141 -----VVTYTPGKPLNTVTT-----LKPSSIGAST-----PSNE 170
DB 707 TQIIQTKGPLPAGTILKLVTSADGKPTIITTTQASGAGTPTIIGTSSVSPSTTKPGTT 766
QY 171 PNLKAENSAVQINLSPTMLENVKCKNFKLAML-IKLACSGSOSPQEMQNVKRLVEQLLD 229
DB 767 TIITIPMSALITQAGATGVTSPEIKSPIITITTKVMTSGTGA-----AKIITAV-- 818
QY 230 AKIAEEFTRL-YVELKSSP-QP-----HLVPFLKSVVALROLNPSQSFIOOCVOQTS 283
DB 819 PKIATGHQOGCVTVLKGAGQPGTILRTVP-----MGVRLVTPVVSVAVKPVP----- 869
QY 284 SDMYIATCTTIVTSPVTTVTVSSQS-----EKSIIVSGATAPRTVSQVLNP 332
DB 870 -TTLVVKGTGTGVTGLGVTGTGVSISLAGAGHSTASLAPITTLGLTIA--TLSSQVNP 926
QY 333 LAGPVGAKAGVTLHSVGPATAGGTT 359
DB 927 TA--ITVSAQOTL-----TAAAGLTT 946

```

```

RESULT 9

```

US-08-244-189-2
; Sequence 2, Application US/08244189
; Patent No. 5380727
; GENERAL INFORMATION:
; APPLICANT: OHKI, Misao
; APPLICANT: KIKUCHI, Kimiko
; APPLICANT: MIYOSHI, Hiroyuki
; APPLICANT: KOTU, Tomoko
; TITLE OF INVENTION: No. 5580727el DNAs, polypeptides Encoded Thereby
; TITLE OF INVENTION: and Methods for Detecting the DNAs and Polypeptides
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,189
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 760-183P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 752 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-244-189-2

Query Match 9.3%; Score 164.5; DB 1; Length 752;
Best Local Similarity 28.8%; Pred. No. 1.1e-06;
Matches 72; Conservative 41; Mismatches 108; Indels 29; Gaps 11;

QY 61 MIVSPQPTVTRAEITSNITSRPAVPANPQTVKICTVPSNSSOLIKKVAATPVKKLAOIGT 120
DB 152 VFTNPQVATYHRAIKITVDGPREPRN-RTEKSTMPDSDPVYKQTSRLTP----- 201
QY 121 TVTIVPKPSSVQSAVAVPTSVTPGKPLNTVTTLKPSL-GASSTP--SNEPNLKAIE 176
DB 202 ---PTMPPTPTQG-APRTSSFPPT---TLTNGTSHSPALNCAAPSPNGFSSSSS 254
QY 177 NSAAVQINISPTM-LENVKKCKNFMILIKIACSSQSPSEMGNVKKVLEQLDAIEM 235
DB 255 SSIAND-QLPACAGARQLSKRELTTLQOF--GNDISPISGERVTLVLGVNSTLTLE 311
QY 236 EFTKRLVELKSSPPHLPFLKKSVAL-ROLLPNSQSFQ---OCVOOTSSDMYIATC 291
DB 312 EFHRSKQEAENTNPLRFVILPELKANLPLLORELLHCARLAKQNPAYLMAQHEQLLDAS 371
QY 292 TTTVTTSPPV 301
DB 372 TSPVDSSELL 381

RESULT 10
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:

; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEBCH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PR1
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-5

Query Match 8.4%; Score 148.5; DB 3; Length 1721;
Best Local Similarity 21.9%; Pred. No. 0.00012;
Matches 82; Conservative 41; Mismatches 183; Indels 69; Gaps 12;

QY 35 TTTIOPANQLPPTVILKSNGLMIVSPQTVTRAEITSNITSRPAVPANPQTVKIC 94
DB 286 TT 345
QY 95 TVPSSSOLIKKVAATPVKKLAOIGTIVTVPKPSSVQSAVAVPTSVTPGKPLNTVT 154
DB 346 TT 398
QY 155 TLKPSLSGASSTPS-----NEPNLKAENSA-----VOINLSPTMDEN-- 192
DB 399 TTAATTTTSETSESVIKPDEMCWLEKNGECEAGATVGVIGKDIENGMATMIPND 458
QY 193 -----VKCKNFMILIKIACSSQSPSEMGNVKKVLEQLDAKIE--AEFTKRLY 242
DB 459 TVHFRFRKVDGNTISVRCRKGAKLEFP-----DRSLDPTIPVAGHNSCSII 508
QY 243 VELKSSPPHLPFLKKSVALROLPLNSQSF-----IQCV-----QOTSSDMYIA 289
DB 509 VGVSQDGKIHVSFGSKDVSLSAPIQPSLEFNEVCDICTAKYGAHSGYQTSADPVT 568
QY 290 TC---TTVTTSPPVTTTVSSSQSEKSIIVGATAPRIVSVQTLNPLAGPVGAKAGVTL 346
DB 569 TPAKPTTTTGAPGQPTTTTGSPSKP--TTTTTKATTTTTTLNPLITTTQKPTTTT 626
QY 347 HSV--GPTAATGCT 359
DB 627 TKVPGKPIATTTTT 641

RESULT 11
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERSEN, VERNY, JONES & BIRKA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vervy, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480,76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-6

Query Match 8.4%; Score 148.5; DB 3; Length 1721;
Best Local Similarity 21.9%; Pred. No. 0.00012;

Matches 82; Conservative 41; Mismatches 183; Indels 69; Gaps 12;

QY 35 TTTIOFANLQLPQGVILKSNGLMLVSPQOTVRAETTSNITSRPVAPNPQTVKIC 94
DB 286 TTTTITT 345
QY 95 TVPSSQLIKKVAATPVKLAQIGTVTVTPKSSVQSAVPTSVYVTPGAPLMTVT 154
DB 346 TTTTITT 398
QY 155 TLKPSISLASSTPS-----NEPNLKAENSA-----VOINLSPMLEN-- 192
DB 399 TTTATTTTTTETESVIRPDCEWLEKNGECCAKATYGVIGKGRLENGMAFTMIRDD 458
QY 193 -----VKCKNFLAMLIKACSSGSSPEMGONVKKLVQILDARIE--AEEFTRLY 242
DB 459 THVFERKVDVGNITISVRCKGAKLEFP-----DRSLDFTLPVAGHNSCII 508
QY 243 VELKSSQPHLVPFLKSVVALRQLPNSQSF-----IQCCY-----QOTSSDVAIA 289
DB 509 VGVSGDKIHVSPGSKDVLISAPIPSELFEVYCDTCAKYGAIHSGYQTSADPVT 568
QY 290 TC-----TTTTTSPVTTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKGVVTL 346
DB 569 TTKAPTTTTTGABQGPPTTTTGSRSKP--TTTTTKATTTTTTILNPIITTTTQKPTTTT 626
QY 347 HSV--GPTAATGTT 359
DB 627 TKVPGKPIATTTTT 641

RESULT 12
US-07-757-022B-14
Sequence 14, Application US/07757022B
Patent No. 6431142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive

CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luanan

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-14

Query Match 8.1%; Score 144; DB 4; Length 941;
Best Local Similarity 20.3%; Pred. No. 0.00012;

Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 APVSAAPRVSSGPHLAPQIVAVKAPNT-----TTIOPANLQLPQGVILI 53
DB 482 APTTKEAPPTTKEPAPPTTKEPAPPTTKEPAPPTTKEPAPPTTKEPAPPTT 538
QY 54 KSNGLMLVSPQOTVRAETTSNITSRPVAPNPQTVKICTVPSQLIKKVAATPVK 113
DB 539 --TKEPSTSDKAPPTPKGTATPTTKEPAPPTTKEPAPPTTKEPAPPTTKEPAPPT 596
QY 114 KLA-----QIGTVTVTPKSSVQSAV--VPTSVTVTPGKPLNVTTLKPSLG 162
DB 597 KPAPELAPPTTCKPTSTSDKPAPTTKEPAPPTTKEPAPPTTKEPAPPTTKEPAPPT 656
QY 163 ASSTP--SNEPNL--KKNL--KKNL--KKNL--KKNL--KKNL--KKNL--KKNL-- 210
DB 657 EVSTPTTKEPTTTHKSDSTPELSAEPYKALENSKEPVGPTTTPAATKPEMTT 716
QY 211 -----SSPEMGONVKKLVQILDARIEAEEFTRLYVELKSSQPHILV 254
DB 717 KDKTTERDLRTPTTETAPAPKKEATTEKTESKITAT-----TQVTSITTOOT 770
QY 255 PFLKSSVALRQLPNSQSFIOCCVQOTSDMVATCTTTVTSPVTVTVSSQSEKSI 314
DB 771 PF-KITTLKTYTLAPK-----VTTTKKITTTTEIMNKKPEETAKPKDRA 812
QY 315 IVSGATAPR 323
DB 813 TNSKATTPK 821

RESULT 13
US-07-757-022B-84
Sequence 84, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-84
Query Match 8.1%; Score 144; DB 4; Length 1022;
Best Local Similarity 20.3%; Pred. No. 0.00014;
Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 AVSAPKPVSSGRRLPAQIVAVKAPNT-----TTIQFANLQLPPTVL 53
DB 563 APTTPKEAPATTPKEPAATTPKETAATTPKGAATPPLKEBAPATTPKPAKELAPTT--- 619
QY 54 KNSGCLMIVSQQVYTRAEITSNITSRPAVNPOTVVICIYVNSSSQLIKKVAATPVK 113
DB 620 --TKETSTSTSDKPAATTPKGAATTPKPAATTPKPAATTPKGAATTPKPAATTPK 677
QY 114 KLA-----QIGTVVTVTPKPSVQSV--VPTSVTVTPGKPLMTVTLKSSIG 162
DB 678 KPAKELATTTTKGPTSTSDKPAATTPKETAATTPKKEPAATTPKPAATTPPTPTPTTS 737
QY 163 ASSTP--SNEPNL--KAENSAVQINLSPT--MLEN-----VKCKNFLAMLIKILACSG 210
DB 738 EVSTPTTKETPTTIHKSPDSETPBELSAEPTPKALENSPKPEGVPTTKTAAATKPEMTTAA 797

QY 211 -----SQSPKGMNKVLEQLDAKIEAEFTKRLYLKSSPQPHLY 254
DB 798 KDKTERDLTPPETTTAAKMTKETATTEKTESKITAT--TQVSTTTODTPT 851
QY 255 PFLKSVVALROLLPNSQSIQOCVQOTSSDMVIACTTIVTSPVTTTVSSQSEKSI 314
DB 852 PF-KITTKITTTIAFK-----VTTTKTITTTETLNKPETAKPDRA 893
QY 315 IVSGATAPR 323
DB 894 TNSKATTPK 902
RESULT 14
US-07-757-022B-74
Sequence 74, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-74
Query Match 8.1%; Score 144; DB 4; Length 1038;
Best Local Similarity 20.3%; Pred. No. 0.00014;
Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 AVSAPKPVSSGRRLPAQIVAVKAPNT-----TTIQFANLQLPPTVL 53

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 38.818 Seconds
(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Perfect score: 562
Sequence: 1 AQPpVLSQpAGIPtGSSSK.....ASPTOKRIKENVtSCFRDE 110

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 segs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*

2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*

3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*

4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*

5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*

6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*

7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*

8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*

9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*

10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*

11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*

12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*

13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*

14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*

15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*

16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*

17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*

18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*

19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*

20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*

21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*

22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*

23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*

24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*

25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*

26: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*

27: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	562	100.0	801	1	PCT-US01-08631-39827
2	562	100.0	852	21	US-09-763-909-2
3	534	95.0	843	1	PCT-US02-29864-410
4	493.5	87.8	865	1	PCT-US02-25829-25
5	302	53.7	685	27	US-60-243-468-1271
6	92.5	16.5	737	1	PCT-US94-01114-16

7	92.5	16.5	1051	1	PCT-US02-30474-1660	Sequence 1660, Ap
8	92.5	16.5	1051	27	US-60-324-631-1665	Sequence 1665, Ap
9	92.5	16.5	1083	1	PCT-US02-30474-3271	Sequence 3271, Ap
10	84.5	15.0	624	27	US-60-389-987-1773	Sequence 1773, Ap
11	84.5	15.0	624	27	US-60-412-418-1773	Sequence 1773, Ap
12	80	14.2	471	16	US-09-248-796-14351	Sequence 14351, Ap
13	80	14.2	471	16	US-09-248-796-14351	Sequence 14351, Ap
14	80	14.2	1398	25	US-60-096-409-14351	Sequence 14351, Ap
15	79.5	14.1	3726	20	US-09-614-150-18633	Sequence 6261, Ap
16	78.5	14.0	622	20	US-09-614-150-18633	Sequence 18633, Ap
17	78.5	14.0	622	20	US-60-161-932-1300	Sequence 7944, Ap
18	78.5	14.0	622	27	US-60-161-932-1300	Sequence 1300, Ap
19	78.5	14.0	622	27	US-60-161-932-1300	Sequence 1300, Ap
20	78.5	14.0	622	27	US-60-173-464-6463	Sequence 8067, Ap
21	78.5	14.0	622	27	US-60-173-464-6463	Sequence 6463, Ap
22	77.5	13.8	622	27	US-60-191-637-7965	Sequence 7965, Ap
23	77.5	13.8	622	27	US-60-191-637-7965	Sequence 7965, Ap
24	77.5	13.8	545	21	US-09-791-537-99180	Sequence 99180, Ap
25	77.5	13.8	545	21	US-09-791-537-99180	Sequence 99180, Ap
26	77.5	13.8	567	19	US-09-514-245-6	Sequence 10506, Ap
27	76.5	13.6	567	19	US-09-514-245-6	Sequence 6, Appl1
28	76.5	13.6	455	16	US-09-270-767-45531	Sequence 4531, Ap
29	76.5	13.6	455	16	US-09-270-767-45531	Sequence 4531, Ap
30	75.5	13.4	3065	1	PCT-US02-25829-19	Sequence 15332, Ap
31	75.5	13.4	441	15	US-09-134-000-6703	Sequence 15, Appl1
32	74.5	13.3	557	15	US-09-134-000-6703	Sequence 6703, Ap
33	74.5	13.3	3085	27	US-60-191-637-18689	Sequence 5061, Ap
34	74.5	13.3	3085	27	US-60-191-637-18689	Sequence 18689, Ap
35	74.5	13.3	3678	27	US-60-167-217-18725	Sequence 14788, Ap
36	74.5	13.3	3678	27	US-60-167-217-18725	Sequence 18725, Ap
37	73.5	13.1	278	26	US-60-171-627-1494	Sequence 1494, Ap
38	73.5	13.1	278	26	US-60-171-627-1494	Sequence 1494, Ap
39	73.5	13.1	278	26	US-10-219-999-62512	Sequence 62512, Ap
40	73	13.0	591	1	PCT-US01-08117-22	Sequence 22, Appl1
41	73	13.0	591	26	US-10-221-625-22	Sequence 22, Appl1
42	73	13.0	591	27	US-60-188-986-22	Sequence 22, Appl1
43	73	13.0	595	18	US-09-488-725A-1831	Sequence 1831, Ap
44	73	13.0	619	18	US-09-488-725A-1830	Sequence 1830, Ap
45	73	13.0	619	21	US-09-791-537-18995	Sequence 18995, Ap

ALIGNMENTS

RESULT 1
PCT-US01-08631-39827
Sequence 39827, Application PC/TUS0108631
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
FILE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-049
CURRENT APPLICATION NUMBER: PCT/US01/08631
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO 39827
LENGTH: 801
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-08631-39827

Query Match 100.0%; Score 562; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 7.6e-53;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQPpVLSQpAGIPtGSSSKQLFLFHYVQPSGNGKQVTTTSHSSTLTITOKGCGATMP 60
DB 443 AQPpVLSQpAGIPtGSSSKQLFLFHYVQPSGNGKQVTTTSHSSTLTITOKGCGATMP 502

QY 61 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 110
Db 503 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 552

RESULT 2

US-09-763-909-2
Sequence 2, Application US/09763909
GENERAL INFORMATION:
APPLICANT: Dikstein, Rivka
TITLE OF INVENTION: A TRANSCRIPTION FACTOR TFIID SUBUNIT,
TITLE OF INVENTION: TAFII105, POLYPEPTIDES, DNA ENCODING THEREFOR AND
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS
FILE REFERENCE: 13005/002001
CURRENT APPLICATION NUMBER: US/09/763, 909
CURRENT FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 852
TYPE: PRT
ORGANISM: Homo sapiens
US-09-763-909-2

Query Match 100.0%; Score 562; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 8,3e-53;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AOPGPVLSOPAGIPTSSSKQLFSLFHVVOOPSGNGEKQVTTSHSSTLTIOKCGQKTPM 60
Db 443 AOPGPVLSOPAGIPTSSSKQLFSLFHVVOOPSGNGEKQVTTSHSSTLTIOKCGQKTPM 502

QY 61 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 110
Db 503 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 552

RESULT 3

PCT-US02-29964-410
Sequence 410, Application PC/TUS0229964
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Ren, Feiyun
APPLICANT: Zhang, Jie
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Xue, Aidong J.
APPLICANT: Wang, Dunrui
APPLICANT: Ghosh, Malabika
APPLICANT: Asundi, Vinod
APPLICANT: Wehrman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Goodrich, Ryle W.
APPLICANT: Weng, Gezh
APPLICANT: Haley-Vicente, Dana
APPLICANT: Dimaec, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 809ACIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/29964
CURRENT FILING DATE: 2002-09-19
PRIOR APPLICATION NUMBER: US 60/323, 739
PRIOR FILING DATE: 2001-09-19
PRIOR APPLICATION NUMBER: US 09/488, 725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552, 317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491, 404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623

PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496, 914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560, 875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515, 126
PRIOR FILING DATE: 2000-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 992
SOFTWARE: PL_FL_genes Version 6.0
SEQ ID NO 410
LENGTH: 843
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-29964-410

Query Match 95.0%; Score 534; DB 1; Length 843;
Best Local Similarity 95.5%; Pred. No. 1.1e-49;
Matches 105; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AOPGPVLSOPAGIPTSSSKQLFSLFHVVOOPSGNGEKQVTTSHSSTLTIOKCGQKTPM 60
Db 485 AOPGPVLSOPAGIPQAVQKQLFSLFHVVOOPSGNGEKQVTTSHSSTLTIOKCGQKTPM 544

QY 61 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 110
Db 545 VNTIIPSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSQFRDE 594

RESULT 4

PCT-US02-25829-25
Sequence 25, Application PC/TUS0225829
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: BAROSSO, Ines
APPLICANT: BAUGHN, Mariah R.
APPLICANT: BECHA, Shanya D.
APPLICANT: BLAKE, Julie J.
APPLICANT: BOROMSKY, Mark L.
APPLICANT: BUREORD, Neil
APPLICANT: DUGGAN, Brendan M.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: EMERLING, Brooke M.
APPLICANT: FORSYTHE, Ian J.
APPLICANT: GIETZEN, Kimberly J.
APPLICANT: GORVAD, Ann E.
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: HAFALIA, April J.A.
APPLICANT: HONCHELL, Cynthia D.
APPLICANT: ISON, Craig H.
APPLICANT: KHAN, Farrah A.
APPLICANT: LAL, Preeti G.
APPLICANT: LEE, Ernestine A.
APPLICANT: LEE, Sally
APPLICANT: LEE, Soo Yeun
APPLICANT: LI, Joana X.
APPLICANT: LU, Dying Alina M.
APPLICANT: LU, Yan
APPLICANT: LEHR-MASON, Patricia M.
APPLICANT: NGUYEN, Daniel B.
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: SPRAGUE, William W.
APPLICANT: TANG, Y. Tom
APPLICANT: THANGAVELU, Kavitha
APPLICANT: THORNTON, Michael
APPLICANT: TRAN, Uyen K.
APPLICANT: WALIA, Narinder K.
APPLICANT: WARREN, Bridget A.
APPLICANT: XU, Yuning
APPLICANT: YAO, Montique G.
APPLICANT: YUE, Henry


```

? APPLICANT: Ren, Feiyan
? APPLICANT: Xue, Aidong J.
? APPLICANT: Zhao, Qing A.
? APPLICANT: Wang, Jian-Rui
? APPLICANT: Wehrman, Tom
? APPLICANT: Zhou, Ping
? APPLICANT: Ghosh, Malabika
? APPLICANT: Wang, Dunrui
? APPLICANT: Ma, Yungqing
? APPLICANT: Asundi, Vinod
? APPLICANT: Wang, Zhiwei
? APPLICANT: Meng, Gezhi
? APPLICANT: Haley-Vicente, Dana
? APPLICANT: Drmanac, Radoje T
? TITLE OF INVENTION: Novel Nucleic Acids and
? FILE REFERENCE: Polypeptides
? CURRENT APPLICATION NUMBER: PCT/US02/30474
? PRIOR FILING DATE: 2002-09-24
? PRIOR APPLICATION NUMBER: US 60/324,631
? PRIOR FILING DATE: 2001-09-24
? PRIOR APPLICATION NUMBER: US 09/488,725
? PRIOR FILING DATE: 2000-01-21
? PRIOR APPLICATION NUMBER: US 09/552,317
? PRIOR FILING DATE: 2000-04-25
? PRIOR APPLICATION NUMBER: PCT/US00/35017
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: US 09/491,404
? PRIOR FILING DATE: 2000-01-25
? PRIOR APPLICATION NUMBER: PCT/US01/02623
? PRIOR FILING DATE: 2001-01-25
? PRIOR APPLICATION NUMBER: US 09/496,914
? PRIOR FILING DATE: 2000-02-03
? PRIOR APPLICATION NUMBER: US 09/560,875
? PRIOR FILING DATE: 2000-04-27
? PRIOR APPLICATION NUMBER: PCT/US01/03800
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 09/515,126
? PRIOR FILING DATE: 2000-02-28
? Remaining prior application data removed - See file wrapper or PALM.
? NUMBER OF SEQ ID NOS: 3476
? SOFTWARE: PL_FL_genes Version 6.0
? SEQ ID NO 1660
? LENGTH: 1051
? TYPE: PRT
? ORGANISM: Homo sapiens
? PCT-US02-30474-1660

Query Match      16.5%; Score 92.5; DB 1; Length 1051;
Best Local Similarity 27.3%; Pred. No. 2.1;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

QY 2 QGPGVLS--QPGIPGSSSKQLEFHVQPGSGNEKQVTTISHSTLTICGQKTM 59
DB 684 QP-PVLSLQPTGVGKQGPPLVIOQPPKGGALIRPPQVLTITPMAALQ----- 736
QY 60 PVTITPT-----SQFPASILKQITLPGNKIL--SLQASPTQXRIKENVTSQFRD 109
DB 737 PHRRILTTPOQIOLNPLQPVVPKPAVLPGRALSAVSAQAAAAQKNKLEGGSGFRD 796
QY 110 E 110
DB 797 D 797

RESULT 8
US-60-324-631-1665
; Sequence 1665, Application US/60324631
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
```

```

? APPLICANT: Zhao, Qing A.
? APPLICANT: Wang, Jian-Rui
? APPLICANT: Wehrman, Tom
? APPLICANT: Zhou, Ping
? APPLICANT: Ghosh, Malabika
? APPLICANT: Wang, Dunrui
? APPLICANT: Ma, Yungqing
? APPLICANT: Asundi, Vinod
? APPLICANT: Wang, Zhiwei
? APPLICANT: Meng, Gezhi
? APPLICANT: Haley-Vicente, Dana
? APPLICANT: Drmanac, Radoje T
? TITLE OF INVENTION: Novel Nucleic Acids and
? FILE REFERENCE: Polypeptides
? CURRENT APPLICATION NUMBER: US/60/324,631
? PRIOR FILING DATE: 2001-09-24
? PRIOR APPLICATION NUMBER: US 09/488,725
? PRIOR FILING DATE: 2000-01-21
? PRIOR APPLICATION NUMBER: US 09/552,317
? PRIOR FILING DATE: 2000-04-25
? PRIOR APPLICATION NUMBER: PCT/US00/35017
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: US 09/491,404
? PRIOR FILING DATE: 2000-01-25
? PRIOR APPLICATION NUMBER: PCT/US01/02623
? PRIOR FILING DATE: 2001-01-25
? PRIOR APPLICATION NUMBER: US 09/496,914
? PRIOR FILING DATE: 2000-02-03
? PRIOR APPLICATION NUMBER: US 09/560,875
? PRIOR FILING DATE: 2000-04-27
? PRIOR APPLICATION NUMBER: PCT/US01/03800
? PRIOR FILING DATE: 2001-02-05
? PRIOR APPLICATION NUMBER: US 09/515,126
? PRIOR FILING DATE: 2000-02-28
? PRIOR APPLICATION NUMBER: US 09/577,409
? PRIOR FILING DATE: 2000-05-18
? PRIOR APPLICATION NUMBER: PCT/US01/04927
? PRIOR FILING DATE: 2001-02-26
? PRIOR APPLICATION NUMBER: US 09/519,705
? PRIOR FILING DATE: 2000-03-07
? PRIOR APPLICATION NUMBER: US 09/574,454
? PRIOR FILING DATE: 2000-05-19
? PRIOR APPLICATION NUMBER: PCT/US01/04941
? PRIOR FILING DATE: 2001-03-05
? PRIOR APPLICATION NUMBER: US 09/540,217
? PRIOR FILING DATE: 2000-03-31
? PRIOR APPLICATION NUMBER: US 09/649,167
? PRIOR FILING DATE: 2000-08-23
? PRIOR APPLICATION NUMBER: PCT/US01/08631
? PRIOR FILING DATE: 2001-03-30
? PRIOR APPLICATION NUMBER: US 09/552,929
? PRIOR FILING DATE: 2000-04-18
? PRIOR APPLICATION NUMBER: US 09/770,160
? PRIOR FILING DATE: 2001-01-26
? PRIOR APPLICATION NUMBER: PCT/US01/08656
? PRIOR FILING DATE: 2001-04-18
? PRIOR APPLICATION NUMBER: US 09/577,408
? PRIOR FILING DATE: 2000-05-18
? PRIOR APPLICATION NUMBER: PCT/US01/14827
? PRIOR FILING DATE: 2001-05-16
? NUMBER OF SEQ ID NOS: 3334
? SOFTWARE: PL_FL_genes Version 6.0
? SEQ ID NO 1665
? LENGTH: 1051
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-60-324-631-1665

Query Match      16.5%; Score 92.5; DB 27; Length 1051;
Best Local Similarity 27.3%; Pred. No. 2.1;
Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;
```

QY 2 QPGVLS--OPAGIPTGSSSKOLFSLFHVVOQPSGNGEKVTTISHSTLTIOKCGQRTM 59
Db 684 QP-VLSTLQTPGVGVGOGPTPLVIOQPKPKALLIRPQVTLTQTPMVALRQ----- 736
QY 60 PVTNIIPT-----SQPPASILKQITLPGNKIL--SLQASPTQKNRIKENTVSCFRD 109
Db 737 PHNIMLTTPQOIQINLPQVPVVKPAVLPGTKALSAVSAQAAAAQKNKLEKPGGGSFRD 796
QY 110 E 110
Db 797 D 797

RESULT 9
PCT-US02-30474-3271
Sequence 3271, Application PC/TUS0230474
GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom
APPLICANT: Zhang, Jie
APPLICANT: Ren, Felyan
APPLICANT: Xue, Aifeng J.
APPLICANT: Zhao, Qidong A.
APPLICANT: Wang, Jian-Rui
APPLICANT: Weinman, Tom
APPLICANT: Zhou, Ping
APPLICANT: Ghosh, Malabika
APPLICANT: Wang, Dunrui
APPLICANT: Ma, Yungling
APPLICANT: Asundi, Vinod
APPLICANT: Wang, Zhilwei
APPLICANT: Wang, Gezhi
APPLICANT: Haley-Vicente, Dana
APPLICANT: Dimanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 810CIP PCT
CURRENT APPLICATION NUMBER: PCT/US02/30474
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: US 60/324,631
PRIOR FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: PCT/US01/03800
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: US 09/515,126
PRIOR FILING DATE: 2000-02-28
Removing prior application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 3476
SOFTWARE: PL_FL-genes Version 6.0
SEQ ID NO 3271
LENGTH: 1083
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1083)
OTHER INFORMATION: Xaa = any amino acid or symbol as shown in the table 8 as set for
PCT-US02-30474-3271

Query Match 16.5%; Score 92.5; DB 1; Length 1083;
Best Local Similarity 27.3%; Pred. No. 2.2;

Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;
QY 2 QPGVLS--OPAGIPTGSSSKOLFSLFHVVOQPSGNGEKVTTISHSTLTIOKCGQRTM 59
Db 716 QP-VLSTLQTPGVGVGOGPTPLVIOQPKPKALLIRPQVTLTQTPMVALRQ----- 768
QY 60 PVTNIIPT-----SQPPASILKQITLPGNKIL--SLQASPTQKNRIKENTVSCFRD 109
Db 769 PHNIMLTTPQOIQINLPQVPVVKPAVLPGTKALSAVSAQAAAAQKNKLEKPGGGSFRD 828
QY 110 E 110
Db 829 D 829

RESULT 10
US-60-389-987-1773
Sequence 1773, Application US/60389987
GENERAL INFORMATION:

APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Wernock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
FILE REFERENCE: 660088.465P2
CURRENT APPLICATION NUMBER: US/60/389,987
CURRENT FILING DATE: 2002-06-17
NUMBER OF SEQ ID NOS: 3025
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 1773
LENGTH: 624
TYPE: PRT
ORGANISM: Homo sapiens
US-60-389-987-1773

Query Match 15.0%; Score 84.5; DB 27; Length 624;
Best Local Similarity 28.9%; Pred. No. 7.9;
Matches 37; Conservative 11; Mismatches 41; Indels 39; Gaps 7;

QY 3 PGPVLSOPAGIP-----TGSSSKOLFSLFHVVOQPSGNGEKVTTISHSTLTIOKCGQ 57
Db 302 PGPASQPGROPPLSPASAGDT-----RGVPQSWGPEAOAASASSPLEALAC-LK 352
QY 58 TNPVTNIIPTSQPPASILKQITLPGNKILSLQASPTQKNRIKEN----- 102
Db 353 GIPNGSSP-SQLPPTS-CSQNPQPGD-----SRQKPELQPRSHSEATREPVLP 403
QY 103 -VTSQCFRD 109
Db 404 GIGSCVRD 411

RESULT 11
US-60-412-418-1773
Sequence 1773, Application US/60412418
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Wernock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
FILE REFERENCE: 660088.465P3
CURRENT APPLICATION NUMBER: US/60/412,418
CURRENT FILING DATE: 2002-09-20
NUMBER OF SEQ ID NOS: 3025

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1773
LENGTH: 624
TYPE: PRT
ORGANISM: Homo sapiens
US-60-412-418-1773

Query Match 15.0%; Score 84.5; DB 27; Length 624;
Best Local Similarity 28.3%; Pred. No. 7.9;
Matches 37; Conservative 11; Mismatches 41; Indels 39; Gaps 7;

3 PGFVLSQPAIGIP-----TSSSSKQLFLFHVVOQPSGNGEKQVTTISHSTLTQKCGOK 57
DB 302 PGFASQGPQRQPLSPATBDT-----RGVQPSMGPEAQAASASSPFLALEAC-LK 352
QY 58 TNPVNTIPTSQFPASILKQITLPGNKILSLQASPTQKNRIKEN----- 102
DB 353 GIPPNSSP-SQLPPTS-CSQNPQPD-----SRSQPELPQPHSHSEATREPVLP 403
QY 103 -VTSQFRD 109
DB 404 GLOSCVRD 411

RESULT 12
US-09-248-796-14351
Sequence 14351, Application US/09248796
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196, 132
CURRENT APPLICATION NUMBER: US/09/248,796
CURRENT FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 28206
SEQ ID NO 14351
LENGTH: 471
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796-14351

Query Match 14.2%; Score 80; DB 16; Length 471;
Best Local Similarity 28.3%; Pred. No. 17;
Matches 28; Conservative 15; Mismatches 48; Indels 8; Gaps 2;

QY 6 VLSQPAIGIPGSSSKQLFLFHVVOQPSGNGEKQVTTISHSTLTQKCGOKTMPVNTII 65
DB 376 VQSOPFG---GFVQPSFQOPQLVSOQRTGPMQAPT-----GSLQAPGTGRPGMGFVSM 427
QY 66 PTQSPFASILKQITLPGNKILSLQASPTQKNRIKENVT 104
DB 428 PTGIGIPGLNMQOHFLPNQLPTSLNLSAMDKNLKENVT 466

RESULT 13
US-60-096-409-14351
Sequence 14351, Application US/60096409A
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: GTC-016P
CURRENT APPLICATION NUMBER: US/60/096,409A
CURRENT FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28206
SEQ ID NO 14351
LENGTH: 471
TYPE: PRT
ORGANISM: Candida albicans
US-60-096-409-14351

Query Match 14.2%; Score 80; DB 27; Length 471;
Best Local Similarity 28.3%; Pred. No. 17;

Matches 28; Conservative 15; Mismatches 48; Indels 8; Gaps 2;
QY 6 VLSQPAIGIPGSSSKQLFLFHVVOQPSGNGEKQVTTISHSTLTQKCGOKTMPVNTII 65
DB 376 VQSOPFG---GFVQPSFQOPQLVSOQRTGPMQAPT-----GSLQAPGTGRPGMGFVSM 427
QY 66 PTQSPFASILKQITLPGNKILSLQASPTQKNRIKENVT 104
DB 428 PTGIGIPGLNMQOHFLPNQLPTSLNLSAMDKNLKENVT 466

RESULT 14
US-10-179-131-6261
Sequence 6261, Application US/10179131
GENERAL INFORMATION:
APPLICANT: HARE, ROBERTA S.
APPLICANT: SHAW, KAREN J.
APPLICANT: SHIMER JR., GEORGE H.
APPLICANT: KESSLER, MARCO
APPLICANT: NOLLING, JORR
APPLICANT: ZENG, QIANDONG
APPLICANT: GREENE, JONATHAN R.
TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
FILE REFERENCE: 2976-4031
CURRENT APPLICATION NUMBER: US/10/179,131
CURRENT FILING DATE: 2002-06-21
NUMBER OF SEQ ID NOS: 10194
SEQ ID NO 6261
LENGTH: 1398
TYPE: PRT
ORGANISM: Candida albicans
US-10-179-131-6261

Query Match 14.2%; Score 80; DB 25; Length 1398;
Best Local Similarity 28.3%; Pred. No. 77;
Matches 28; Conservative 15; Mismatches 48; Indels 8; Gaps 2;

QY 6 VLSQPAIGIPGSSSKQLFLFHVVOQPSGNGEKQVTTISHSTLTQKCGOKTMPVNTII 65
DB 376 VQSOPFG---GFVQPSFQOPQLVSOQRTGPMQAPT-----GSLQAPGTGRPGMGFVSM 427
QY 66 PTQSPFASILKQITLPGNKILSLQASPTQKNRIKENVT 104
DB 428 PTGIGIPGLNMQOHFLPNQLPTSLNLSAMDKNLKENVT 466

RESULT 15
US-09-614-150-18633
Sequence 18633, Application US/09614150
GENERAL INFORMATION:
APPLICANT: Venter, J. Craig
APPLICANT: et al.
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
FILE REFERENCE: CL000728
CURRENT APPLICATION NUMBER: US/09/614,150
CURRENT FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/157,832
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: 60/160,191
PRIOR FILING DATE: 1999-10-19
PRIOR APPLICATION NUMBER: 60/161,932
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/164,769
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/173,383
PRIOR FILING DATE: 1999-12-28
PRIOR APPLICATION NUMBER: 60/175,693
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: 60/184,831
PRIOR FILING DATE: 2000-02-24

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:00:19 ; Search time 4.47537 seconds
(without alignments)
2076.603 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Perfect score: 562
Sequence: 1 AAGPVLSPAGIPGSSSK.....ASPTOKRIKENVTSCEFRDE 110

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 429898 seqs, 84487048 residues

Total number of hits satisfying chosen parameters: 429898

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/PCIT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	18.0	925	US-09-724-676-63928	Sequence 63928, A
2	101	18.0	925	US-09-724-676-63928	Sequence 63928, A
3	101	18.0	957	US-09-724-676-63931	Sequence 63931, A
4	101	18.0	957	US-09-724-676-63931	Sequence 63931, A
5	101	18.0	1021	US-09-724-676-63922	Sequence 63922, A
6	101	18.0	1021	US-09-724-676-63922	Sequence 63922, A
7	101	18.0	1053	US-09-724-676-63925	Sequence 63925, A
8	101	18.0	1053	US-09-724-676-63925	Sequence 63925, A
9	93	16.5	679	US-09-724-676-63929	Sequence 63929, A
10	93	16.5	679	US-09-724-676-63929	Sequence 63929, A
11	93	16.5	711	US-09-724-676-63920	Sequence 63920, A
12	93	16.5	711	US-09-724-676-63920	Sequence 63920, A
13	93	16.5	775	US-09-724-676-63923	Sequence 63923, A
14	93	16.5	775	US-09-724-676-63923	Sequence 63923, A
15	93	16.5	807	US-09-724-676-63926	Sequence 63926, A
16	93	16.5	807	US-09-724-676-63926	Sequence 63926, A
17	84.5	15.0	616	US-10-245-882-264	Sequence 264, App
18	84.5	15.0	616	US-10-245-882-264	Sequence 264, App
19	84.5	15.0	709	US-09-724-676-63927	Sequence 63927, A
20	84.5	15.0	709	US-09-724-676-63927	Sequence 63927, A
21	84.5	15.0	741	US-09-724-676-63930	Sequence 63930, A
22	84.5	15.0	741	US-09-724-676-63930	Sequence 63930, A
23	84.5	15.0	805	US-09-724-676-63921	Sequence 63921, A
24	84.5	15.0	805	US-09-724-676-63921	Sequence 63921, A
25	84.5	15.0	837	US-09-724-676-63924	Sequence 63924, A
26	84.5	15.0	837	US-09-724-676-63924	Sequence 63924, A

27	84.5	15.0	1030	1	PCT-US02-29560-263	Sequence 263, App
28	84.5	15.0	1030	6	US-10-245-882-263	Sequence 263, App
29	75.5	13.4	441	5	US-09-134-000C-6703	Sequence 6703, App
30	75.5	13.4	441	5	US-09-134-000C-6703	Sequence 6703, App
31	75.5	13.4	557	5	US-09-134-000C-5061	Sequence 5061, App
32	75.5	13.4	557	5	US-09-134-000C-5061	Sequence 5061, App
33	74	13.2	617	6	US-10-144-779-555	Sequence 555, App
34	72.5	12.9	412	6	US-10-092-411A-3773	Sequence 3773, App
35	71	12.6	528	1	PCT-US02-04915-14	Sequence 14, App
36	70.5	12.5	128	1	PCT-US02-33727-7803	Sequence 7803, App
37	70.5	12.5	128	6	US-10-057-498-7803	Sequence 7803, App
38	70	12.5	410	5	US-09-751-708A-48	Sequence 48, App
39	69.5	12.4	1532	6	US-10-218-140-1418	Sequence 1418, App
40	68.5	12.2	186	5	US-09-724-676-87265	Sequence 87265, A
41	68.5	12.2	186	5	US-09-724-676-87265	Sequence 87265, A
42	68.5	12.2	213	5	US-09-724-676-87264	Sequence 87264, A
43	68.5	12.2	213	5	US-09-724-676-87264	Sequence 87264, A
44	68.5	12.2	579	5	US-09-724-676-87260	Sequence 87260, A
45	68.5	12.2	579	5	US-09-724-676-87260	Sequence 87260, A

ALIGNMENTS

RESULT 1
US-09-724-676-63928
; Sequence 63928, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63928
; LENGTH: 925
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63928

Query Match 18.0%; Score 101; DB 5; Length 925;
Best Local Similarity 29.2%; Pred. No. 0.14;
Matches 38; Conservative 16; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGVLSPPAGIPG-----SSSKQLFSLFHYQSGGSEKRYT-----TISHSSLT 50
Db 551 QPPPTSOATTALAAVLSSS-----VORTAGKTAATVTSALDPVLSLTQPIQVG 601
QY 51 IQKCGKQTMPV-----NTIPTSQPPASILKQITLPGNKIL---SLQASPTOKNRK 100
Db 602 VGKGGQPIPLVYIQPPKRGALIQNLPQPVVPAVPLPGTKALSAVSAQAAAAQKNK 661
QY 101 ENWTSCEFRDE 110
Db 662 EPGGGSFRDD 671

RESULT 2
US-09-724-676A-63928
; Sequence 63928, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63928
; LENGTH: 925
; TYPE: PRT
; ORGANISM: Homo sapiens

QY 51 IQKCGQKTMPIV-----NTIIPTSQFPFASILKQITLPGNKIL---SLQASPTQKNRIK 100
Db 698 VGGGQGPPLVIOQPPKRGALIQLNLPQVPVVKPAVLPGTKALSAVSAQAQAAAAQKNKRLK 757
QY 101 ENVTSCFRDE 110
Db 758 EPGGGSFRDD 767

RESULT 7
US-09-724-676-63925
; Sequence 63925, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63925
; LENGTH: 1053
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63925

Query Match 18.0%; Score 101; DB 5; Length 1053;
Best Local Similarity 29.2%; Pred. No. 0.17;
Matches 38; Conservative 16; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGVLSQPAIGPIG---SSSKQLFSLFHVVOQPSGNEKQVT-----TISHSSTLT 50
Db 679 QPPPTSQATATLTAFLVLSSS-----VQRTAGKTAATVTSALQPPVLSLTQPTQVG 729
QY 51 IQKCGQKTMPIV-----NTIIPTSQFPFASILKQITLPGNKIL---SLQASPTQKNRIK 100
Db 730 VGGGQGPPLVIOQPPKRGALIQLNLPQVPVVKPAVLPGTKALSAVSAQAQAAAAQKNKRLK 789
QY 101 ENVTSCFRDE 110
Db 790 EPGGGSFRDD 799

RESULT 8
US-09-724-676A-63925
; Sequence 63925, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63925
; LENGTH: 1053
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63925

Query Match 18.0%; Score 101; DB 5; Length 1053;
Best Local Similarity 29.2%; Pred. No. 0.17;
Matches 38; Conservative 16; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGVLSQPAIGPIG---SSSKQLFSLFHVVOQPSGNEKQVT-----TISHSSTLT 50
Db 679 QPPPTSQATATLTAFLVLSSS-----VQRTAGKTAATVTSALQPPVLSLTQPTQVG 729
QY 51 IQKCGQKTMPIV-----NTIIPTSQFPFASILKQITLPGNKIL---SLQASPTQKNRIK 100
Db 730 VGGGQGPPLVIOQPPKRGALIQLNLPQVPVVKPAVLPGTKALSAVSAQAQAAAAQKNKRLK 789
QY 101 ENVTSCFRDE 110

Db 790 EPGGGSFRDD 799

RESULT 9
US-09-724-676-63929
; Sequence 63929, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63929
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63929

Query Match 16.5%; Score 93; DB 5; Length 679;
Best Local Similarity 28.9%; Pred. No. 0.6;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGVLSQPAIGPIG---SSSKQLFSLFHVVOQPSGNEKQVT-----TISHSSTLT 50
Db 551 QPPPTSQATATLTAFLVLSSS-----VQRTAGKTAATVTSALQPPVLSLTQPTQVG 601
QY 51 IQKCGQKTMPIV-----NTIIPTSQFPFASILKQITLPGNKIL---SLQASPTQKNRIK 100
Db 602 VGGGQGPPLVIOQPPKRGALIQLNLPQVPVVKPAVLPGTKALSAVSAQAQAAAAQKNKRLK 661
QY 101 ENVTSCFR 108
Db 662 EPGGGSFR 669

RESULT 10
US-09-724-676A-63929
; Sequence 63929, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63929
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63929

Query Match 16.5%; Score 93; DB 5; Length 679;
Best Local Similarity 28.9%; Pred. No. 0.6;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGVLSQPAIGPIG---SSSKQLFSLFHVVOQPSGNEKQVT-----TISHSSTLT 50
Db 551 QPPPTSQATATLTAFLVLSSS-----VQRTAGKTAATVTSALQPPVLSLTQPTQVG 601
QY 51 IQKCGQKTMPIV-----NTIIPTSQFPFASILKQITLPGNKIL---SLQASPTQKNRIK 100
Db 602 VGGGQGPPLVIOQPPKRGALIQLNLPQVPVVKPAVLPGTKALSAVSAQAQAAAAQKNKRLK 661
QY 101 ENVTSCFR 108
Db 662 EPGGGSFR 669

RESULT 11

```
US-09-724-676-63920
; Sequence 63920, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63920
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-63920

Query Match
Best Local Similarity 28.9%; Pred. No. 0.63;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGPVLSQPAIGPTG---SSSKQLFSLFHVVOQPSGNGEKQVT-----TISHSSTLT 50
Db 583 QPPPTSQATTTALTAVLSSS-----VQRTAGKTATVTSALQPPVLSLTQPTQVG 633
QY 51 IQKCGQKTMPIV-----NTIIPTSQFPASILKQITLPGNKIL---SIQASPTOKNRK 100
Db 634 VGGKGGPTPLVIOQPPKPGALIQLNPLQVPVVKPAVLPGTKALSAVSAQAAAAQKNK 693
QY 101 ENVTSCFR 108
Db 694 EPGGGSFR 701

RESULT 12
US-09-724-676A-63920
; Sequence 63920, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63920
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676A-63920

Query Match
Best Local Similarity 28.9%; Pred. No. 0.63;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGPVLSQPAIGPTG---SSSKQLFSLFHVVOQPSGNGEKQVT-----TISHSSTLT 50
Db 583 QPPPTSQATTTALTAVLSSS-----VQRTAGKTATVTSALQPPVLSLTQPTQVG 633
QY 51 IQKCGQKTMPIV-----NTIIPTSQFPASILKQITLPGNKIL---SIQASPTOKNRK 100
Db 634 VGGKGGPTPLVIOQPPKPGALIQLNPLQVPVVKPAVLPGTKALSAVSAQAAAAQKNK 693
QY 101 ENVTSCFR 108
Db 694 EPGGGSFR 701

RESULT 13
US-09-724-676-63923
; Sequence 63923, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
```

```
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63923
; LENGTH: 775
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-63923

Query Match
Best Local Similarity 28.9%; Pred. No. 0.69;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGPVLSQPAIGPTG---SSSKQLFSLFHVVOQPSGNGEKQVT-----TISHSSTLT 50
Db 647 QPPPTSQATTTALTAVLSSS-----VQRTAGKTATVTSALQPPVLSLTQPTQVG 697
QY 51 IQKCGQKTMPIV-----NTIIPTSQFPASILKQITLPGNKIL---SIQASPTOKNRK 100
Db 698 VGGKGGPTPLVIOQPPKPGALIQLNPLQVPVVKPAVLPGTKALSAVSAQAAAAQKNK 757
QY 101 ENVTSCFR 108
Db 758 EPGGGSFR 765

RESULT 14
US-09-724-676A-63923
; Sequence 63923, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63923
; LENGTH: 775
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676A-63923

Query Match
Best Local Similarity 28.9%; Pred. No. 0.69;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

QY 2 QPGPVLSQPAIGPTG---SSSKQLFSLFHVVOQPSGNGEKQVT-----TISHSSTLT 50
Db 647 QPPPTSQATTTALTAVLSSS-----VQRTAGKTATVTSALQPPVLSLTQPTQVG 697
QY 51 IQKCGQKTMPIV-----NTIIPTSQFPASILKQITLPGNKIL---SIQASPTOKNRK 100
Db 698 VGGKGGPTPLVIOQPPKPGALIQLNPLQVPVVKPAVLPGTKALSAVSAQAAAAQKNK 757
QY 101 ENVTSCFR 108
Db 758 EPGGGSFR 765

RESULT 15
US-09-724-676-63926
; Sequence 63926, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
```

SEQ ID NO 63926
LENGTH: 807
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-63926

Query Match 16.5%; Score 93; DB 5; Length 807;
Best Local Similarity 28.9%; Pred. No. 0.72;
Matches 37; Conservative 15; Mismatches 46; Indels 30; Gaps 5;

OY 2 QPGPVLISQPGACIPTC-----SSSKOLFSLFHVVOQPSGNEKQVT-----TISHSTLT 50
Db 679 QPPPTSQATTALTAVVLSS-----VQRTAGTAATVTSALQPPVLSLTQPTQVG 729
OY 51 IQKCGQKTMV-----NTIIPTSQPPASILKQITLPGNKIL---SIQASPTQKNRIK 100
Db 730 VQKQGPPLVLIQPPKPGALIQLNPLQPVVVKPAVLPGTRKLSAVSAQAAAAQKNKIK 789
OY 101 ENVTSCFR 108
Db 790 EPGGGSFR 797

Search completed: February 16, 2003, 22:04:28
Job time : 6.47537 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:55:19 ; Search time 3.72163 Seconds
(without alignments)
2841.441 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Perfect score: 562
Sequence: 1 AOPGPVLSOPAGIPGTSSSK.....ASPTQKNRIKENYTSCEFRDE 110

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80.5	14.3	410	2 T50075	probable nicotinat
2	79.5	14.1	3759	2 A35085	trithorax protein
3	78	13.9	346	2 C75572	hypothetical prote
4	76.5	13.6	636	2 S63131	probable membrane
5	76	13.5	403	2 T39846	probable nuclear p
6	74	13.5	434	2 T43197	nucleoporin homolo
7	74	13.2	615	2 JC7576	transcription fact
8	74	13.0	1777	2 T43369	hypothetical prote
9	73	13.0	619	2 A43361	ETS-related transc
10	73	13.0	1466	2 A36426	SPB2 protein - yea
11	72.5	12.9	444	1 C64226	trigger factor Mg2
12	72.5	12.9	545	2 B53309	probable pheromone
13	71.5	12.7	642	2 T39607	fork head protein
14	71.5	12.7	979	2 A35913	regulatory factor
15	70.5	12.5	865	2 AC1966	hypothetical prote
16	70	12.5	218	2 S37658	drought-induced pr
17	70	12.5	441	2 T23461	hypothetical prote
18	70	12.5	688	2 F64111	DNA-directed DNA p
19	70	12.5	1484	2 T42632	breast cancer tumo
20	69.5	12.4	412	1 B70125	hemolysin homolog
21	69.5	12.4	767	2 T19690	hypothetical prote
22	69.5	12.4	1366	2 B86292	F7H2.12 protein -
23	69	12.3	97	2 C22848	hypothetical ORF-2
24	69	12.3	578	2 T48795	origin recognition
25	68.5	12.2	663	2 T40493	hmf-3/forkhead tra
26	68	12.1	379	2 E71296	probable spermidin
27	68	12.1	1858	2 T18273	1 phosphatidylinos
28	67.5	12.0	768	2 T37601	probable transcript
29	67	11.9	366	2 A49076	transcription fact

30	67	11.9	5376	2 T42215	zonadhesin - mouse
31	66.5	11.8	237	1 S73990	ribosomal protein
32	66.5	11.8	343	2 A90654	ATP-binding compon
33	66.5	11.8	343	2 A85505	ATP-binding compon
34	66.5	11.8	510	2 H84834	En/Spn-like transp
35	66.5	11.8	536	2 T17217	hypothetical prote
36	66.5	11.8	1194	2 E96624	hypothetical prote
37	66	11.7	248	2 F96655	hypothetical prote
38	66	11.7	552	2 T24869	probable transcript
39	66	11.7	753	2 T24869	hypothetical prote
40	66	11.7	1129	2 T19779	hypothetical prote
41	66	11.7	2531	2 A46019	Notch-1 protein -
42	65.5	11.7	247	2 C86204	hypothetical prote
43	65.5	11.7	343	2 C64744	probable ABC-type
44	65.5	11.7	343	2 AC0533	probable ABC trans
45	65.5	11.7	466	2 S45419	hypothetical prote

ALIGNMENTS

RESULT 1
T50075
probable nicotinate phosphoribosyltransferase [imported] - fission yeast (Schizosacch
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: T50075
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.
submitted to the EMBL Data Library, December 1999
A:Reference number: Z25034
A:Accession: T50075
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Residues: 1-410 <MCD>
A:Molecule type: DNA
A:Cross-references: EMBL:AL133557; PIRN:CAB62416.1; GSPDB:GN00066; SPDB:SPAC1486.06
A:Experimental source: strain 972h(-); cosmid c1486
C:Genetics:
A:Gene: SPDB:SPAC1486.06
A:Map position: 1
C:Superfamily: nicotinate phosphoribosyltransferase

Query Match 14.3%, Score 80.5; DB 2; Length 410;
Best Local Similarity 24.8%; Pred. No. 2.3;
Matches 29; Conservative 21; Mismatches 38; Indels 29; Gaps 5;

QY 15 TGSSSKOLFSEHVQPSG-----GNEKQVTTISHSTLT----- 51
DB 273 TANSADLANVFGVQDSGAEYIEKVYKXISGVDPSTKTVIHSDALNVRDIELY 332
QY 52 --QKCGQRT-MPVNTIIPTSQFPASIIKQITLPGNKILSL-QASPTQKNRIKENV 103
DB 333 KYCEKGIKSAFGICITNL-TSDFQKVSNPSEVKSMNIVILFSAEGTKAKAISIDI 388

RESULT 2
A35085
trithorax protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Sep-1999
C:Accession: A35085; A38240
R:Mazo, A.M.; Huang, D.H.; Mozer, B.A.; David, I.B.
Proc. Natl. Acad. Sci. U.S.A. 87, 2112-2116, 1990
A:Title: The trithorax gene, a trans-acting regulator of the bithorax complex in Dros
A:Reference number: A35085; MUID:90192757; PMID:2107543
A:Accession: A35085
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3759 <MAZ>
R:Mazo, A.M.
submitted to Genbank, January 1990
A:Reference number: A38240
A:Accession: A38240
A:Molecule type: mRNA

A:Residues: 1-2361, 'Y', 2363-2397, 'N', 2399-2405, 'N', 2407-2411, 'N', 2413-3759 <MA2>
A:Cross-references: GB:M31017; NID:g158817; PID:g158818

C:Genetics:

A:Gene: FlyBase:trix

A:Cross-references: FlyBase:FBgn0003862

C:Superfamily: Drosophila trichorax protein

C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match

Best Local Similarity 14.1%; Score 79.5; DB 2; Length 3759;
Matches 30; Conservative 24; Mismatches 50; Indels 61; Gaps 5;

3 PGVLSOPACIPITGSSSKOLFSLFHVVOQPSGNGEKVY-----40

Db 2986 PKVTISQOR-IPAOYQOQOQOQOMHIPPQOQPLQOQOYVQPSMDITLAEAPVVOQ 3044

QY 41 -----TTTSSSTLTIOCKGQKTPVNTIPTSOPFPAISIKQITLPENKI 86

Db 3045 FVMEPOLAEQOELANRQVHSTSSSSSSNCILPTVNVNMQOAPSTTSSSTTRPTNRV 3104

QY 87 LSLQ-----ASPTQKNRIKE-----NVTSCF 107

Db 3105 LPMQORQEPAPLSNECPVSSPTPKPEVQPIIHQMTASVSKCY 3149

RESULT 3

C75572 hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: C75572

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MIMD:20036986; PMID:10567266

A:Accession: C75572

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-346 <MH1>

A:Cross-references: GB:AE001864; GB:AE000513; NID:g6457659; PIDN:AAF09603.1; PID:g645766

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0008

A:Map position: 1

Query Match

Best Local Similarity 13.9%; Score 78; DB 2; Length 346;
Matches 24; Conservative 24; Mismatches 45; Indels 16; Gaps 2;

QY 9 QPAGIPGSSSKOLFSLFHVVOQPSGNGEKVYTTISSTLTIOCKGQK-----HSTLTIO 52

Db 175 RPSEVQATAPSRVILNQVTRVTSVPDLDPCEERAPLIALDSGRVVEPTLHPASVSQ 234

QY 53 KCGQKTPVNTIPTSOPFPAISIKQITLPENKIISQASPTQKNRIKE 101

Db 235 RLDITGLPVTLPVVLGAPPASLRVOSQTLQPTRVVVAPELLGLRL 283

RESULT 4

S63131 probable membrane protein YNL176c - yeast (Saccharomyces cerevisiae)

M:Alternate names: hypothetical protein N161

C:Species: Saccharomyces cerevisiae

C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002

C:Accession: S63131

R:Obermaier, B.; Piravandi, E.; Rinke, M.; Dondley, H.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S63122

A:Accession: S63131

A:Molecule type: DNA

A:Residues: 1-636 <OBE>

A:Cross-references: EMBL:Z71452; NID:g1302153; PID:e239549; PID:g1302154; GSPDB:GN000

A:Experimental source: strain 5288C

C:Genetics:

A:Gene: MIPS:YNL176c

A:Cross-references: SGD:S0005120

A:Map position: 14L

C:Keywords: transmembrane protein

F:464-480/Domain: transmembrane #status predicted <TM>

Query Match

Best Local Similarity 13.6%; Score 76.5; DB 2; Length 636;
Matches 23; Conservative 12; Mismatches 34; Indels 7; Gaps 2;

QY 7 LSOPAGIPGSSSKOLFSLFHVVOQPSGNGEKVYTTISSTLTIOCKGQKTPVNTIIP 66

Db 296 ISVPTSSSVSSSSSKVPS-----NRPSSSSSDDTTAVSYTTFQSLQSTTS--SSIPP 348

QY 67 TSOPFPAISIKQITLP 82

Db 349 TTOPPTSTTISTSPDP 364

RESULT 5

C739846 probable nuclear pore protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39846

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.

submitted to the EMBL Data Library, February 1998

A:Reference number: Z21884

A:Accession: T39846

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-403 <LYN>

A:Cross-references: EMBL:AL021839; PIDN:CAA17069.1; GSPDB:GN00067; SPDB:SPBC1967.15

A:Experimental source: strain 972n; cosmid c1967

C:Genetics:

A:Gene: SPDB:SPBC1967.15

A:Map position: 2

A:Introns: 6/1

Query Match

Best Local Similarity 13.5%; Score 76; DB 2; Length 403;
Matches 30; Conservative 14; Mismatches 49; Indels 18; Gaps 3;

QY 11 AGIPGSSSKOLF--SLFHVVOQPSGNGEKVYTTISSTLTIOCKGQK-----58

Db 50 AGGPTGSSSAPPFGNSIFGRTQOQPTTSFSNTTTNAGOSTVFGONASRTGNSNTOPLFS 109

QY 59 -----MPVNTIPTSOPFPAISIKQITLPENKIISQASPTQKNRIKENY 103

Db 110 WSTVNNPTKEVDETNTATIPSSLLSSGISPNATVSNQYQAPAPVSEEGY 160

RESULT 6

T43197 nucleoporin homolog - fission yeast (Schizosaccharomyces pombe) (fragment)

C:Species: Schizosaccharomyces pombe

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T43197

R:Toshioaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.

DNA Res. 4, 363-369, 1997

A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.

A:Reference number: Z17323; MIMD:98162722; PMID:9501991

A:Accession: T43197

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-434 <YOS>

A:Cross-references: EMBL:DB9269; NID:g1749745; PIDN:BAAL3930.1; PID:g1749746

A:Experimental source: strain PR745

Query Match

13.5%; Score 76; DB 2; Length 434;

Best Local Similarity 27.0%; Pred. No. 6.9;
Matches 30; Conservative 14; Mismatches 49; Indels 18; Gaps

```
OY 11 AGITGSSKOLF--SLFHYVOOPSGGNGEQTITSHSLTIQKGGQT----- 58
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 AGPGTSSAPPFNSIFGTQOOPPFSENTTTPNAPQSTVFQGNASRTGNSNTOPLFS 140
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 59 -----MPVNTIIPYSGPPASILKQITL-EGNKILSIQASPTQGNKRIENY 103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 MSTVNNPPTKPYDETINATIPSSLLSSGISINATVNSNAQGPAPQPEVEEQY 191
```

```

RESULT 7
JC7576
transcription factor Elf-1, type 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-2001 #sequence=revision 30-Jun-2001 #text-change 30-Jun-2001
C:Accession: JC7576
R:Nishiyama, C.; Takahashi, K.; Nishiyama, M.; Okumura, K.; Ra, C.; Ohtake, Y.; Yokota, N.;
Biosci. Biotechnol. Biochem. 64, 2601-2607, 2000
A:Title: Splice isoforms of transcription factor Elf-1 affecting its regulatory function
A:Reference number: JC7576; PMID: 21077473; PMID:11210123
A:Contents: Mast cell line, RBL-2H3
A:Accession: JC7576
A:Molecule type: mRNA
A:Residues: 1-615 <NTS>
A:Cross-references: DDBJ:AB030215
C:Comment: This protein, as a key transcription factor for immune-related genes, has the
C:Gene: Elf-1

```

Query Match	Similarity	13.2%	Score 74	DB 2	Length 615
Best Local	Similarity	30.9%	Pred. No. 17		
Matches	30	Conservative	11	Mismatches	38
				Indels	18
				Gaps	5
Qy	2	QPGPLYS--QPGIGIPITGSSSKOLFSLFHHVQOPSGNGEKQVNTISHSSPLTIQKQGQKTM	59		
		:		:	
		:		:	
Db	362	QPSFVLKRVYQPSQADYPT---OLFRTVHHVQVQVAIPEEATIT---STMOEEAAGSSVQ	414		
Qy	60	PVNTIIPTSQEPFASILKQITLPGKKIL---SLQASP	93		
		:			
Db	415	GIRTIQASTQVP-----VVSFPGKQQLHTVTLQTVP	445		

```

RESULT 8
T34369
hypothetical protein t19d12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision.29-Oct-1999 #text_change.29-Oct-1999
C:Accession: t34369
R:Favelllo, A.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid t19d12.
A:Reference number: 221513
A:Accession: t34369
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1777 <FAV>
A:Cross-references: EMBL:U01263; PRDN:AAC24428.1; GSPDS:GN00020; CESP:t19d12.1
A:Experimental source: strain Bristol N2; clone t19d12
C:Genetics:
A:Gene: CESP:t19d12.1
A:Map position: 2
A:Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681/1

```

Query Match	13.2%	Score 74;	DB 2;	Length 1777;
Best Local Similarity	27.2%	Pred. No. 58;		
Matches	34;	Conservative	18;	Mismatches 41; Indels 32; Gaps 6.

Oy	1	AQPGVYLQAPACIPFGSSSKQLFLFHHVQDPSGGNGEKQVTTTISSTITLTKG-----	55
Db	928	AQPAOSSSTSAATVTVSS-----QSTRSSPAQDSSSTPACSSSTVTVQOSSSFOSP	977

```

Oy 56 OKTAPV--NTTPTSOEPPASI-----LKOTTLGG--NKILSLASPTOKNRK 100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db . 978 GSTTGSSSTVPESTQAPSSISGGPTTQICPNQGVIFRGVGVIIEMLPASTQGNALNA 1037
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 101 --ENV 103
      |||
Db 1038 FVENV 1042

```

RESULT 9
A43361
Ets-related transcription factor Elf-1 - human
N:Alternate names: E74-like factor Elf-1
C:Species: Homo sapiens (man)
C:Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 07-May-1999
C:Accession: A43361; A42122
R:Leiden, J.M.; Wang, C.Y.; Petryniak, B.; Markovitz, D.M.; Nabel, G.J.; Thompson, C.
J. Virol. 66, 5890-5897, 1992
A:Title: A novel Ets-related transcription factor, Elf-1, binds to human immunodeficiency
A:Reference number: A43361; MUID:92407982; PMID:1527846
A:Accession: A43361
A:Molecule type: mRNA
A:Residues: 1-619 <LEI>
A:Cross-references: GB:M8282
R:Thompson, C.B.; Wang, C.Y.; Ho, I.C.; Bohjanen, P.R.; Petryniak, B.; June, C.H.; M
Mol. Cell. Biol. 12, 1043-1053, 1992
A:Title: cis-acting sequences required for inducible interleukin-2 enhancer function
A:Reference number: A42122; MUID:92186836; PMID:1545787
A:Accession: A42122
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 204-282, 'G', 284-289 <THO>
A:Experimental source: T-cells
A:Note: sequence extracted from NCBI backbone (NCBI:86286)

```

A:Gene: GDB:ELF1
A:Cross-references: GDB:131648
A:Map position: 1p36-1p36
C:Superfamily: ets DNA-binding domain homology
C:Keywords: DNA binding; transcription regulation
F:210-290/Domain: ets DNA-binding domain homology <ETS>

Query Match      13.0%  Score 73;  DB 2;  Length 619;
Best Local Similarity 28.9%  Pred. No. 21;
Matches 37;  Conservative 11;  Mismatches 50;  Indels 30;  Gaps 6;

```

```

Oy      1 AOPGVLS--OPAGIPTSSSKOLEFHHVQ---PSGNEKOVYTIHSHSTIQRK 54
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      362 AQPSEVLATVQPTQSPLYT---QLFRIVHVQPAQNAPG--EAARSTMDLTIASSVQ 416

Oy      55 GOKTPPVNTIIITSPFPASTIKOILT-----PGRKLISLQASPIQR- 96
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      417 SIRIQAFQTVPPVVSPNQDLHTVLTQVPLTVIASDPSAGTSQRTILNIPBSQP 476

```

QY	97 -NRKENV	103
	:	
Db	477 MTVLKENV	484

RESULT 10

SPA2 protein - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: protein L1209; protein YL021w
 C:Species: *Saccharomyces cerevisiae*

C/Accession: A36426; S64769; S69386
R/Gehrung, S.; Snyder, M.
J. Cell Biol. 111, 1451-1464, 1990

A:Title: The SPA2 gene of *Saccharomyces cerevisiae* is important for pheromone-induced
A:Reference number: A36426; MUID:91009481; PMID:2211820
A:Accession: A36426
A:Molecule type: DNA

A:Cross-references: GB:X53731; NID:g4519; PID:CAA37763.1; PID:g4520
R:Goffeau, A.; Purnelle, B.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64761
A:Accession: S64761
A:Molecule type: DNA
A:Residues: 1-1466 <GOR>
A:Cross-references: EMBL:Z73126; NID:g1360195; PID:e245456; PID:g1360196; MIPS:YLL021w
A:Note: experimental_source strain S288C
R:Purnelle, B.; Goffeau, A.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of 32 kb on the left arm of yeast chromosome XII reveals 14
mily and a new ABC transporter homologous to the human multidrug resistance protein.
A:Reference number: S69380
A:Accession: S69380
A:Molecule type: DNA
A:Residues: 1-1466 <PUS>
A:Cross-references: EMBL:X97560; NID:g1297003; PID:e238713; PID:g1297010
C:Genetics:
A:Gene: SGD:SPA2
A:Cross-references: SGD:S0003944; MIPS:YLL021w
A:Map position: 12L
C:Keywords: coiled coil

Query Match 13.0%; Score 73; DB 2; Length 1466;
Best Local Similarity 23.1%; Pred. No. 58;
Matches 28; Conservative 27; Mismatches 32; Indels 34; Gaps 6;

OY 5 PVL-SQAPGIPGS---SSKQLFSLF--HVQOPSGGNE-----KQVTTISHST 48
DB 600 PIRPSSNGVPTTSKRKPGSTGLSLMIDSSIAKSHKDNKYSPILAVTSASASAS 659
OY 49 LTIQKCGQKTMPT---VNTIIPISQPPPSILKQITLPGKILSLQASPTQKNRIKENT 104
DB 660 SMISEIPLKTLPPQAKIGVIPPSE-----NQVNPRIKIENTEDNKRSDIT 705
OY 105 S 105
DB 706 N 706

RESULT 11

C64226
triglycer factor MG238 (similarity) - Mycoplasma genitalium
N:Contents: peptidylprolyl isomerase (EC 5.2.1.8), ribosome-bound
C:Species: Mycoplasma genitalium
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C:Accession: C64226
R:Fraser, C.M.; Nguyen, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhmann, J.; Nguyen, D.; Usterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346; PMID:7569993
A:Accession: C64226
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-444 <TIK>
A:Cross-references: GB:U03701; GB:I43967; NID:g1045915; PID:g1045927; TIGR:MG238
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: trigger factor; BKR-type peptidylprolyl isomerase homology
C:Keywords: cis-trans-isomerase
E:170-214/Domain: BKR-type peptidylprolyl isomerase homology <PPI>

Query Match 12.9%; Score 72.5; DB 1; Length 444;
Best Local Similarity 26.1%; Pred. No. 16;
Matches 31; Conservative 18; Mismatches 31; Indels 39; Gaps 6;

OY 10 PGIPTSSSKQLFSLFHVQOPSGGNEKQVTTISHSTLTQKCGQKTMPTVNTIPTSQ 69
DB 267 PPGQFSTSLKAFKSFYHILME---NKKQETILQENN---QKIQPLT-TTKTLP--- 314

OY 70 FPPASITKQITLPGKILSLQASPTQKNRI-----KENVT 104
DB 315 FLPEALIK---LEARNILKIQSQAEQYKIPFEKLASNTITTELDNRNIKEKENV 370

RESULT 12

B53309
Probable pheromone binding protein prgz - Enterococcus faecalis plasmid pCF10
N:Alternate names: pheromone responsive gene 2 protein
C:Species: Enterococcus faecalis
C>Date: 13-Sep-1995 #sequence_revision 13-Sep-1995 #text_change 20-Aug-1999
C:Accession: B53309
R:Ruhfel, R.E.; Mantas, D.A.; Dunny, G.M.
J. Bacteriol. 175, 5253-5259, 1993
A:Title: Cloning and characterization of a region of the Enterococcus faecalis conj
A:Reference number: A53309; MUID:93352432; PMID:8349565
A:Accession: B53309
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-545 <RUH>
A:Cross-references: GB:L14285; NID:g309660; PIDN:AAA25554.1; PID:g309662
C:Genetics:
A:Gene: prgz
A:Genome: plasmid
C:Superfamily: dipeptide transport protein

Query Match 12.9%; Score 72.5; DB 2; Length 545;
Best Local Similarity 26.1%; Pred. No. 20;
Matches 23; Conservative 15; Mismatches 29; Indels 21; Gaps 3;

OY 15 TGSSSKQLFSLFHVQOPSGGNEKQVTTISHSTLTQKCGQKTMPTVNTIPTSQPPAS 74
DB 135 TVSPNVELFSAIRNAKIASGKQAK-----DLAVKSIQKTELTLEVPYPPF---- 183
OY 75 ILKQITLPGKILSLQASPTQKNRIKE 101
DB 184 -----TDLSTLTATYTPVQQAIIKE 202

RESULT 13

T39607
fork head protein type transcription factor - fission yeast (Schizosaccharomyces pombe
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39607
R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrett, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21866
A:Accession: T39607
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-642 <LYN>
A:Cross-references: EMBL:AL023554; PIDN:CAA19034.1; GSPDB:GN00067; SPDB:SPBC1665.15C
A:Experimental source: strain 97zh-; cosmid c1665
C:Genetics:
A:Gene: SPDB:SPBC1665.15C
A:Map position: 2
A:introns: 171/2

Query Match 12.7%; Score 71.5; DB 2; Length 642;
Best Local Similarity 28.3%; Pred. No. 31;
Matches 28; Conservative 15; Mismatches 39; Indels 17; Gaps 4;

OY 17 SSSKQLFSLFHVQOP-----SGGNEKQV---TTISHSTLTQKCGQKTMPTVNT 63
DB 480 SSPKTIQDLQGLDVLVNGFEGISSWRESVYVNTLRSSVDSPTVNLANSNSKSSPV-A 538
OY 64 IIPISQPPASITKQITLPGKILSLQASPTQKNRIKEN 102
DB 539 VQVSTLPPQASANKQAKEMESK---MSNSPTQSKRIEEN 574

RESULT 14

A35913
 regulatory factor X - human
 C:Species: Homo sapiens (man)
 C:Date: 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change 05-Nov-1999
 C:Accession: A35913
 R:Reilly, W.; Herrero-Sanchez, C.; Kober, M.; Silacci, P.; Berle, C.; Barras, E.; Fey, S.;
 Genes Dev. 4, 1528-1540, 1990
 A:Title: MHC class II regulatory factor RFX has a novel DNA-binding domain and a function
 A:Reference number: A35913; MUID:1071581; PMID:2253877
 A:Accession: A35913
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-979 <REI>
 A:Cross-references: GB:X58964; NID:g311362; PIDN:CAA41730.1; PID:g33568
 C:Keywords: DNA binding; transcription regulation

Query Match 12.78; Score 71.5; DB 2; Length 979;
 Best Local Similarity 22.88; Pred. No. 51;

Matches 34; Conservative 17; Mismatches 43; Indels 55; Gaps 6;

```

QY 1 AOPGPVLSOPAGIPT-----GSSSKOLFSLFHYVQO 31
      | | | | |
Db 80 AVFAP--SQPTGAPTPSPAPQYIVTVYSEGAMASSETVSASPESTASQTGVPPTQVYQO 137
      | | | | |
QY 32 PSNGNERQVTTIS-----HSTLTIOKCGQKTPVNTIIPTSQFPASILKQITLPGNK 85
      | | | | |
Db 138 VQGTOQLVLTSTVQAKPGHVSPIQL-----TNIQVPOALPTQRLVYVQSAAPGSK 188
      | | | | |
QY 86 --ILSLDA-----SPTQKNRIKENVTS 105
      | | | | |
Db 189 GGOVSLTVHGTQGVHSPEDSPVOANSSS 217
      | | | | |

```

RESULT 15

AC1966
 hypothetical protein alr1278 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AC1966
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Triguera,
 Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AC1966
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-865 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BAF73235.1; PID:g17130625; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr1278
 C:Superfamily: Synechocystis hypothetical protein sl10350

Query Match 12.58; Score 70.5; DB 2; Length 865;
 Best Local Similarity 29.58; Pred. No. 55;

Matches 33; Conservative 11; Mismatches 43; Indels 25; Gaps 5;

```

QY 5 PVLSPAGIPTGSSSKOLFSLFHYVQFSGGNERQVTTISHS-----STLPIQ 52
      | | | | |
Db 84 PVLSPKSPETPTSE---FS--PLTPSSAVNLGQTALIGSVQNNTSANSPPVPTATIP 138
      | | | | |
QY 53 KCGQKTPVNTIIPTSQFPASILKQITLPGNKILSLQASPTQKNRIKENVT 104
      | | | | |
Db 139 QSVQNMSTNSPV---FPTAKILPDVSP---TKLWASPLTRKQEQKDEVT 182
      | | | | |

```

Search completed: February 16, 2003, 22:00:55
 Job time : 6.72163 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 2.11991 Seconds

(Without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552

Sequence: 1 AOPGPVLSQAPAGIPTGSSSK.....ASPTQKRIKENVTSCFRDE 110

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	562	100.0	801	T2DT_HUMAN	O92750 homo sapien
2	92.5	16.5	1083	T2D3_HUMAN	O00268 homo sapien
3	80.5	14.3	410	NPT1_SCHPO	Q0ULK3 schizosacch
4	79.5	14.1	3726	TRX_DROME	P20659 drosophila
5	76.5	13.6	612	ELF1_MOUSE	O60775 mus musculu
6	76.5	13.6	636	YNR6_YEAST	P53882 saccharomyc
7	73	13.0	619	ELF1_HUMAN	P32519 homo sapien
8	73	13.0	1466	SPA2_YEAST	P23201 saccharomyc
9	72.5	12.9	444	TIG_MYCE	P47480 mycoplasma
10	71.5	12.7	979	REF1_HUMAN	P22670 homo sapien
11	71	12.6	528	PODX_HUMAN	O00592 homo sapien
12	71	12.5	1531	NFT5_HUMAN	O94916 homo sapien
13	70	12.5	688	DP3X_HAEN	P43746 haemophilus
14	69	12.3	551	SMA4_MOUSE	P97471 mus musculu
15	68.5	12.2	239	RGSK_MOUSE	O99q21 mus musculu
16	68.5	12.2	551	RN27_MOUSE	O99q21 mus musculu
17	68.5	12.2	663	SEPI_SCHPO	O43058 schizosacch
18	68.5	12.2	2300	CYAA_NEUCR	O01631 neosporea
19	68	12.1	552	SMA4_RAT	O70437 rattus norv
20	68	12.1	1858	P3K2_DICDI	P54674 dictyostell
21	67.5	12.0	2442	CBP_HUMAN	O92750 homo sapien
22	67	11.9	366	T2AA_DROME	P52654 drosophila
23	67	11.9	552	SMA4_PIG	O9qk99 sus scrofa
24	67	11.8	5376	ZAN_MOUSE	O88799 mus musculu
25	66.5	11.8	237	EL23_MYCPN	P75578 mycoplasma
26	66.5	11.8	239	CUS8_HUMAN	P98505 homo sapien
27	66	11.7	194	RS7_METVA	P14037 methanococc
28	66	11.7	154	SMA4_HUMAN	O13485 homo sapien
29	66	11.7	2531	NTC1_MOUSE	O01705 mus musculu
30	65.5	11.7	343	ABC_ECOLI	P30705 escherichia
31	65.5	11.7	466	YB16_YEAST	P38177 saccharomyc
32	65.5	11.7	628	ABFA_ASPNG	P42234 aspergillus
33	65.5	11.7	983	Y144_HUMAN	O14157 homo sapien

34	65	11.6	284	1	SIC1_YEAST	P38634 saccharomyc
35	65	11.6	448	1	YAP1_CHICK	P46936 gallus galli
36	65	11.6	545	1	AGAL_ASPNG	P28351 aspergillus
37	65	11.6	896	1	EP15_HUMAN	P42566 homo sapien
38	65	11.6	1228	1	ECM_HUMAN	O13201 homo sapien
39	65	11.6	1273	1	WEB1_YEAST	P38968 saccharomyc
40	64.5	11.5	2441	1	CBP_MOUSE	P45481 mus musculu
41	64	11.4	199	1	SOXE_SULAC	O53765 sulfolobus
42	64	11.4	234	1	TNEA_CAPI	P13296 capra hircu
43	64	11.4	322	1	YD67_MYCPN	P75414 mycoplasma
44	64	11.4	326	1	VS09_ROTBT	P30210 bovine tota
45	64	11.4	677	1	VGP_EBORS	O89853 ebola virus

ALIGNMENTS

RESULT 1
T2DT_HUMAN STANDARD: PRT; 801 AA.
AC Q92750;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-2002 (Rel. 41, Last annotation update)
DE Transcription Initiation factor TFIIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (Fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=97011146; PubMed=8858156;
RA Diestein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIIID subunit related to
hTAFII130.";
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIIID THAT MAY FUNCTION AS
A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIIID IS A
MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING
PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM
HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nucleus.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED
CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC EMBL: Y09321; CA70499.1; -
DR Genew: HGNC:11538; TAF4B.
DR MIM: 601689; -
DR InterPro: IPR003894; TAF_hom.
DR SMART: SM00549; TAFH; 1.
KW Transcription regulation; Nuclear protein.
FT NON_TER 1
SQ SEQUENCE 801 AA; 85658 MW; D12B4933FEA49CD2 CRC64;
Query Match 100.0%; Score 562; DB 1; Length 801;
Best Local Similarity 100.0%; Pred. No. 3 5e-48;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AOPGPVLSQAPAGIPTGSSSKQLFSLFHVVOOPSGGENKQVTTISHSTLTIOKCGOKTTP 60
|||||

DB 443 ACPGVLSPAGIPGSSSKOLFSLFHVQOPSGNGKQVTTISHSSTLTITQKCGKTM 502

QY 61 VNTIIPSPASILKQITLPGNKILSLQASPTQKRIKENVTSRDE 110

DB 503 VNTIIPSPASILKQITLPGNKILSLQASPTQKRIKENVTSRDE 552

RESULT 2

T2D3_HUMAN

ID T2D3_HUMAN STANDARD; PRT; 1083 AA.

AC 000268; 099721; 09BX42; 09BR40;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUN-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)

DE (TAFII135) (TAFII-130) (TAFII130).

GN TAF4 OR TAF4A OR TAF2C1 OR TAFII135 OR TAFII130.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-97336072; PubMed-9192867;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Scavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey O.P., Bird C.P., Bates K.N., Beard L.M., Beare D.M., Beasley J., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cooley V.E., Collier R.E., Connor R.E., Cordy N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Gramam D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Leharasaino M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh S.A., Mistry S.L., Mooney M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Philimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;

RA The DNA sequence and comparative analysis of human chromosome 20.*;

RL Nature 414:865-871 (2001).

RN [3]

RP SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.

RX MEDLINE-97098442; PubMed-8942982;

RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;

RT Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII30 and hTAFII100.*;

RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616 (1996).

CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION BY THE AF-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.

CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: Y11354; CAA72189.1; -

DR EMBL: AL137077; CAC36006.1; -

DR EMBL: AL109911; CAC2312.2; -

DR EMBL: U75308; AAC50901.1; -

DR TRANSFAC: T02328; -

DR Genew; HGNC: 11537; TAF4.

DR MIM: 601796; -

DR InterPro: IPR003894; TAF_hom.

DR SMART: SM00549; TAFH.1.

KW Transcription regulation; Nuclear protein.

FT DOMAIN 39 42 POLY-HIS.

FT DOMAIN 52 57 POLY-ALA.

FT DOMAIN 98 101 POLY-GLY.

FT DOMAIN 142 148 POLY-ALA.

FT DOMAIN 268 275 POLY-PRO.

FT DOMAIN 331 337 POLY-ALA.

FT DOMAIN 680 683 POLY-PRO.

FT DOMAIN 808 813 POLY-ALA.

FT DOMAIN 828 831 POLY-ASP.

FT CONFLICT 105 117 PCPSPRPRLVPA -> GNGLLQQRGGRES (IN REF. 3).

FT CONFLICT 136 136 A -> S (IN REF. 2).

FT CONFLICT 185 185 G -> GPG (IN REF. 2).

FT CONFLICT 233 264 MISSING (IN REF. 3).

FT CONFLICT 293 293 P -> L (IN REF. 3).

SO SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CMC64;

Query Match 16.58; Score 92.5; DB 1; Length 1083;

Best Local Similarity 27.38; Pred. No. 0.2;

Matches 33; Conservative 19; Mismatches 50; Indels 19; Gaps 5;

QY 2 QPGVLS-OPAGIPGSSSKOLFSLFHVQOPSGNGKQVTTISHSSTLTITQKCGKTM 59

DB 716 QP-VLSLTPGVGKQGPRLVQOPRKEALIRPOVTLTPVWALQ----- 768

QY 60 PVNTIIPSPASILKQITLPGNKIL--SLQASPTQKRIKENVTSRDE 109

DB 769 PNRIMLTTPQQLQNLNLPVVPVPRVLPCTALSAVSAQAAAKNKKIKERGGSGFRD 828

QY 110 E 110

DB 829 D 829

RESULT 3

NPTL_SCHPO

ID NPTL_SCHPO STANDARD; PRT; 410 AA.

AC 090TK3;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable nicotinate phosphoribosyltransferase (EC 2.4.2.11)

DE (NAPRTase).

GN SPAC1486.06.

OS Schizosaccharomyces pombe (Pislon yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.

OC NCBI_TaxID=4896;

OX [1]

RP SEQUENCE FROM N.A.

RX STRAIN=972;

RL MEDLINE-21848401; PubMed-11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA Jones K., Jones L., Jones M., Leather S., McDonald S., Melan J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Mblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tavey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Wolckart G., Aert R., Roben J., Grymonprez B.,
 RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Filtz C., Holzer E., Moestl D., Odell H.,
 RA Botzpm K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambolt R., Punelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leclaire V., Mottier S.,
 RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "the genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- CATALYTIC ACTIVITY: Nicotinate D-ribonucleotide + diphosphate -
 CC nicotinate + 5-phospho-alpha-D-ribose 1-diphosphate.
 CC -1- SIMILARITY: NAD biosynthesis; nicotinamide to NADN; second step.
 CC -1- SIMILARITY: BELONGS TO THE NADPHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AL133357; CAB62416.1; -
 KW Pyridine nucleotide biosynthesis; Transferase; Glycosyltransferase.
 SO SEQUENCE 410 AA; 46694 MW; E4230ADPD9ABA6077 CMC64;
 CC -----
 QY Query Match 14.3%; Score 80.5; DB 1; Length 410;
 Db Best Local Similarity 24.8%; Pred. NO. 0.98;
 Matches 29; Conservative 21; Mismatches 38; Indels 29; Gaps 5;
 QY 15 TGSSSKOLFSLFHVQGPSG-----GNEKQVTTISHSTLT----- 51
 Db 273 TANSADLIANVFGVRODSGCAEYIEKVXHKYSIGVDPSTKVTHSDALNVDCRIELY 332
 QY 52 ---QKSGQKT-MPVNTIITPSOPPASILIKOTITLPGNKILST-QASPTOKNRKENV 103
 Db 333 KYCEKCGIKSAFGICITNL-TSDFOKVNPSPEVKRPMNIYIKLFSAGTKRAVTSDDI 388
 RESULT 4
 TRX_DROME STANDARD: PRT: 3726 AA.
 ID TRX_DROME
 AC P20659; Q27255; Q27327;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trithorax protein.
 GN TRX.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90192757; PubMed=2107543;

RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;
 RT "The trithorax gene, a trans-acting regulator of the bithorax complex
 RT in Drosophila, encodes a protein with zinc-binding domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
 RX MEDLINE=95009521; PubMed=7924996;
 RA Sedkov Y., Tiliib S., Mizrokh L., Mazo A.;
 RT "The bithorax complex is regulated by trithorax earlier during
 RT Drosophila embryogenesis than is the Antennapedia complex, correlating
 RT with a bithorax-like expression pattern of distinct early trithorax
 RT transcripts.";
 RL Development 120:1907-1917(1994).
 [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN-Oregon-R;
 RC MEDLINE=96100387; PubMed=8555104;
 RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
 RT "Conservation of structure and expression of the trithorax gene
 RT between Drosophila virilis and Drosophila melanogaster.";
 RL Mech. Dev. 53:113-122(1995).
 [4]
 RP CHARACTERIZATION.
 RX MEDLINE=95047388; PubMed=7958911;
 RA Kuzin B., Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
 RT "The Drosophila trithorax gene encodes a chromosomal protein and
 RT directly regulates the region-specific homeotic gene fork head.";
 RL Genes Dev. 8:2478-2490(1994).
 CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
 CC WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
 CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
 CC ZINC.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M31617; AAA29025.1; -
 DR EMBL; Z50152; CAA90514.1; -
 DR EMBL; Z50152; CAA90513.1; -
 DR EMBL; Z31725; CAA83516.1; -
 DR EMBL; Z31725; CAA83515.1; -
 DR PIR; A35085; A35085.
 DR HSSP; P20393; 1A6Y.
 DR TRANSFAC; T00850; -
 DR FLYBase; FBgn0003862; trx.
 DR InterPro; IPR003889; FYRIC_N.
 DR InterPro; IPR003888; FYRIC_C.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00628; PHD; 3.
 DR Pfam; PF00856; SET; 1.
 DR SMART; SM00542; FYRC; 1.
 DR SMART; SM00541; FYRN; 1.
 DR SMART; SM00249; PHD; 4.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00184; RING; 3.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS00280; SET; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 4.

DR PROSITE; PS50016; 2F, PHD_2; 3.
 KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW Nuclear protein; Developmental protein; Activator;
 KW Alternative splicing.
 FT 2N_FING 1266 1347 PHD-TYPE 1.
 FT 2N_FING 1348 1393 PHD-TYPE 2.
 FT 2N_FING 1421 1482 PHD-TYPE 3.
 FT 2N_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
 FT 2N_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
 FT DOMAIN 3599 3708 SET.
 FT DOMAIN 512 516 POLY-SER.
 FT DOMAIN 565 570 POLY-ASP.
 FT DOMAIN 661 664 POLY-SER.
 FT DOMAIN 905 910 POLY-SER.
 FT DOMAIN 1576 1582 POLY-GLN.
 FT DOMAIN 2298 3027 GLN-RICH.
 FT DOMAIN 3032 3040 POLY-SER.
 FT DOMAIN 3181 3184 POLY-GLN.
 FT DOMAIN 3220 3225 POLY-GLU.
 FT VARSPLIC 1 368 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 2023 2025 P -> PWLSPFLKFLGSLTHGGLLWLLGCVVRLKQCG
 FT CONFLICT 2341 2341 (IN REF. 1).
 FT CONFLICT 2392 2392 R -> S (IN REF. 1).
 FT CONFLICT 2392 2392 G -> S (IN REF. 1).
 SQ SEQUENCE 3726 AA; 400575 MM; D275650763D1CF5 CRC64;
 Query Match 14.18; Score 79.5; DB 1; Length 3726;
 Best Local Similarity 18.2%; Pred. No. 17;
 Matches 30; Conservative 24; Mismatches 50; Indels 61; Gaps 5;
 QY 3 PGPLSPAGIPGTSSSKOL-FSLPHVVOGPSGNEKOV----- 40
 DB 2953 PKYTISQQR-IPACTQOOQLQAOAMHIFQOQOPLOQOOVVOVOPSPMTITLAEPPVQSO 3011
 QY 41 -----TTTISHSTLTITQKCGKTMPEVNTITFTSQPPASILKQITLPQNKI 86
 DB 3012 FVMEPALQEOELANRYQHFSTSSSSSSNCLEPTVYVNMQOQAPSTSSSTRPNNRV 3071
 QY 87 LSLQ-----ASPTQKNRIKE-----NVTSCF 107
 DB 3072 LPMQOQREPAPLSNECPVSSPTPPKPEQPIIHQMTSASVSKCY 3116
 RESULT 5
 ELFL_MOUSE
 ID ELFL_MOUSE STANDARD; PRT; 612 AA.
 AC Q60775;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ETS-related transcription factor Elf-1 (E74-like factor 1).
 GN ELFL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Fibroblast;
 RX MEDLINE=96257231; PubMed=8666284;
 RA Davis J.N., Roussee M.F.;
 RT "Cloning and expression of the murine Elf-1 cDNA."
 RL Gene 171:265-269(1996).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT APPEARS TO BE REQUIRED FOR THE
 CC T-CELL-RECEPTOR-MEDIATED TRANS ACTIVATION OF HIV-2 GENE
 CC EXPRESSION. BINDS SPECIFICALLY TO TWO PURINE-RICH MOTIFS IN THE
 CC HIV-2 ENHANCER. ELF-1 BINDS TO THE UNDERPHOSPHORYLATED FORM OF RB.
 CC MAY INTERACT WITH OTHER TRANSCRIPTION FACTORS IN ORDER TO REGULATE
 CC SPECIFIC GENES (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY FOUND IN HEMATOPOIETIC CELLS.
 CC DETECTED IN OTHER CELL TYPES SUCH AS FIBROBLASTS.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U19617; AAB17097.1; -.
 DR HSSP; P28324; IBC8.
 DR TRANSFAC; T05012; -.
 DR MGD; MGI:107180; Elf1.
 DR InterPro; IPR000418; Ets.
 DR InterPro; IPR002341; HSF_ETs.
 DR Pfam; PF00178; Ets; 1.
 DR PRINTS; PR00454; ETSDOMAIN.
 DR SMART; SM00413; ETS; 1.
 DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
 KW Nuclear protein; Transcription regulation; Activator; DNA-binding.
 FT DOMAIN 75 80 POLY-ASP.
 FT DOMAIN 208 290 ETS-DOMAIN.
 SQ SEQUENCE 612 AA; 66221 MM; 442F4C85142B31F0 CRC64;
 Query Match 13.6%; Score 76.5; DB 1; Length 612;
 Best Local Similarity 24.7%; Pred. No. 3.9;
 Matches 38; Conservative 22; Mismatches 37; Indels 57; Gaps 9;
 QY 2 QPGVYLS-QPAGIPGTSSSKOLESLRHVVO-----QPSGNEKOVTT 42
 DB 363 QPSEVLETVQPSQAPVPT---QLERTVHVVOQVAVPEASTMGEANSSVPSIRF 419
 QY 43 ISHSSTLTIT-----QKCGKTMPEVNTITFTSQPPASILKQITLPQNKI 77
 DB 420 IQASTQVYVYSPNQDLHYTVPLTVIVASIDPSSGAGSKFTLQITPSSQ--PMTVLK 477
 QY 78 QITL-----PQNKILSLQASPTQKNRI-KENYTS 105
 DB 478 ENVMLOQSKRGSP--SIVLSPTQVQVLTFSNVGS 509
 RESULT 6
 YNR6_YEAST
 ID YNR6_YEAST STANDARD; PRT; 636 AA.
 AC P53882;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 67.4 kDa protein in RPS3-PSD1 intergenic region.
 GN YNL176C OR N1661.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycas.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Obermaier B., Piravandi E., Rinke M., Domdey H.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z71452; CAA96068.1; -.
 DR SGD; S0005120; YNL176C.
 DR Hypothetical protein; Transmembrane.
 FT DOMAIN 94 98 POLY-SER.

FT DOMAIN 131 135 POLY-SER.
 FT DOMAIN 301 309 POLY-SER.
 FT DOMAIN 317 322 POLY-SER.
 FT TRANSMEM 457 477 POTENTIAL.
 SO SEQUENCE 636 AA; 67356 MW; 4027EF48DEA5E287 CRC64;

Query Match 13.6%; Score 76.5; DB 1; Length 636;
 Best Local Similarity 30.3%; Pred. No. 4.1;
 Matches 23; Conservative 12; Mismatches 34; Indels 7; Gaps 2;

OY 7 LSOPGIPITGSSSKQLFHFVVOQPSGNGEKQVTTISHSSTLTIOKGGQKTPVNTIIP 66
 Db 296 ISVPTSSSVSSSSKVPKPS-----NRPSSSSSDDTTSAVSTYVQSLQSTTS--SSIP 348
 OY 67 TSOEPPASILKQITLP 82
 Db 349 TTOPPSTSTISPIIP 364

RESULT 7
 ELFL_HUMAN STANDARD: PRT: 619 AA.

AC P32519: O9UDEL.
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ETS-related transcription factor Elf-1 (E74-like factor 1).
 GN ELFL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92407982; PubMed-1527846;
 RA Leiden J.M., Wang C.Y., Petryniak B., Markovitz D.M., Nabel G.J.,
 Thompson C.B.;
 RT "A novel Ets-related transcription factor, Elf-1, binds to human
 immunodeficiency virus type 2 regulatory elements that are required
 for inducible trans activation in T cells.";
 RL J. Virol. 66:5890-5897(1992).
 RN [2]
 RP SEQUENCE OF 204-289 FROM N.A.
 RX MEDLINE-92186836; PubMed-1545787;
 RA Thompson C.B., Wang C.Y., Ho I.C., Bohjanen P.R., Petryniak B.,
 June C.H., Miesfeldt S., Zhang L., Nabel G.J., Karpinski B.;
 RT "Cis-acting sequences required for inducible interleukin-2 enhancer
 function bind a novel Ets-related protein, Elf-1.";
 RL Mol. Cell. Biol. 12:1043-1053(1992).
 RN [3]
 RP BINDING TO RB.
 RX MEDLINE-93262492; PubMed-8493578;
 RA Wang C.Y., Petryniak B., Thompson C.B., Kaelin W.G., Leiden J.M.;
 RT "Regulation of the Ets-related transcription factor Elf-1 by binding
 to the retinoblastoma protein.";
 RL Science 260:1330-1335(1993).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT APPEARS TO BE REQUIRED FOR THE
 T-CELL-RECEPTOR-MEDIATED TRANS ACTIVATION OF HIV-2 GENE
 EXPRESSION. BINDS SPECIFICALLY TO TWO PURINE-RICH MOTIFS IN THE
 HIV-2 ENHANCER. ELF-1 BINDS TO THE UNDERPHOSPHORYLATED FORM OF RB.
 MAY INTERACT WITH OTHER TRANSCRIPTION FACTORS IN ORDER TO REGULATE
 SPECIFIC GENES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M82882; -, NOT_ANNOTATED_CDS.
 DR PIR; A43361; A43361.
 DR HSSP; P28324; 1BC8.
 DR TRANSFAC; T01113; -.
 DR Genew; HGNC:3316; ELFL.
 DR MIM; 189973; -.

DR InterPro; IPR000418; Ets.
 DR InterPro; IPR002341; HSF_ETS.
 DR Pfam; PF00178; Ets; 1.
 DR PRINTS; PR00454; ETSDOMAIN.
 DR SMART; SM00413; ETS; 1.

DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
 KW Nuclear protein; Transcription regulation; Activator; DNA-binding.

FT DNA_BIND 75 80 POLY-ASP.
 FT CONFLICT 283 283 ETS-DOMAIN.
 FT CONFLICT 283 283 Q -> G (IN REF. 2).
 SQ SEQUENCE 619 AA; 67455 MW; AB0B41B2964A66EF CRC64;

Query Match 13.0%; Score 73; DB 1; Length 619;
 Best Local Similarity 28.9%; Pred. No. 8.9;
 Matches 37; Conservative 11; Mismatches 50; Indels 30; Gaps 6;

OY 1 AOPGVLVS--OPAGIPITGSSSKQLFHFVVOQ---PSGNGEKQVTTISHSSTLTIOKC 54
 Db 362 AQPSEVLRTVPTQSPYPT---QLFRVHVAVPVQAVPEG--EAARTSYMODETLNSSVQ 416
 OY 55 GQKTPVNTIIPTSOEPASILKQITL-----PQNKLSLQASPTQK- 96
 Db 417 SIRTQAPQVQVNVVSPRNOQLHTVLTQVPLTVIASTPSAGTSGOKFIQAIPSSOP 476
 OY 97 -NRKENV 103
 Db 477 MTVLKENV 484

RESULT 8
 SPA2_YEAST

ID SPA2_YEAST STANDARD: PRT: 1466 AA.
 AC P23201.
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SPA2 protein.
 GN SPA2 OR PEAL OR YIL021W OR L1209.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-S288C;
 RX MEDLINE-91009481; PubMed-2211820;
 RA Gehring S., Snyder M.;
 RT "The SPA2 gene of Saccharomyces cerevisiae is important for
 phenomene-induced morphogenesis and efficient mating.";
 RL J. Cell Biol. 111:1451-1464(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-S288C;
 RX STRAIN-S288C;
 RA Purnelle B., Goffeau A.;
 RA Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: INVOLVED IN PHEROMONE-INDUCED MORPHOGENESIS AND
 CC -1- FUNCTION: INVOLVED IN PHEROMONE-INDUCED MORPHOGENESIS AND
 CC -1- SUBCELLULAR LOCATION: LOCALIZES A SHARP PATCH AT THE SHMOO TIP
 CC (MATING PROJECTION) WHICH IS THE SITE OF POLARIZED CELL GROWTH.
 CC -1- SIMILARITY: TO PROTEINS THAT FORM COILED-COIL STRUCTURES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch.)
 CC -----
 DR EMBL: X53731; CAA37763.1; -
 DR EMBL: X97560; CAA66170.1; -
 DR EMBL: 273126; CAA97469.1; -
 DR PIR: A36426; A36426.
 DR SGD: S0003944; SPR2.
 KW Cell shape: Repeat; Coiled coil.
 FT DOMAIN 286 388 COILED COIL (POTENTIAL).
 FT 1087 818 1087 25 x 9 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 818 825 1.
 FT REPEAT 825 834 2.
 FT REPEAT 835 843 3.
 FT REPEAT 860 868 4.
 FT REPEAT 875 883 5.
 FT REPEAT 884 892 6.
 FT REPEAT 893 901 7.
 FT REPEAT 902 910 8.
 FT REPEAT 911 919 9.
 FT REPEAT 920 928 10.
 FT REPEAT 929 937 11.
 FT REPEAT 938 946 12.
 FT REPEAT 947 953 13.
 FT REPEAT 954 961 14.
 FT REPEAT 962 970 15.
 FT REPEAT 971 979 16.
 FT REPEAT 980 988 17.
 FT REPEAT 989 997 18.
 FT REPEAT 998 1006 19.
 FT REPEAT 1007 1015 20.
 FT REPEAT 1036 1044 21.
 FT REPEAT 1045 1053 22.
 FT REPEAT 1054 1062 23.
 FT REPEAT 1072 1080 24.
 FT REPEAT 1081 1087 25.
 FT REPEAT 1169 1189
 FT DOMAIN 1275 1302 COILED COIL (POTENTIAL).
 FT SEQUENCE 1466 AA: 163142 MW: 288616152382C89 CRC64;
 SQ
 Query Match 13.0%; Score 73; DB 1; Length 1466;
 Best Local Similarity 23.1%; Pred. No. 24;
 Matches 28; Conservative 27; Mismatches 32; Indels 34; Gaps 6;
 QY 5 PVL--SOPACIPGSG--SSKOLFSLF--HYVOQPSGNGE-----KQVTTISHST 48
 DB 600 PIRPSSNGVPTTSRKPSGTGLFLMIDSSIAKNSHKEKDNDKRYVPIKAVTSASNSAS 659
 QY 49 LTIQKCGQKMP---VNTIIPTSOPPASILKQITLPKSKILSLASPTOKRIKENVT 104
 DB 660 SNISEIPKLTLPQAKIGVIPSE-----NQVPNIKIENTEEDNKRSDIT 705
 QY 105 S 105
 DB 706 N 706
 RESULT 9
 TIG_MYCGE
 ID TIG_MYCGE STANDARD; PRT; 444 AA.
 AC P47480;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trigger factor (TF).
 GN TIG OR MG238.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;

RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Goodyne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fitchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
 RA Nguyen D.T., Uitterlinden T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Luster T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 RN [2]
 RP SEQUENCE OF 141-216 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 sequencing";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. ACTS AS A CHAPERONE BY
 MAINTAINING THE NEWLY SYNTHESIZED PROTEIN IN AN OPEN CONFORMATION
 (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY. TIG SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch.)
 CC -----
 DR EMBL: U39703; AAC71459.1; -
 DR EMBL: U01772; AAD10591.1; -
 DR TIG: MG238;
 DR InterPro: IPR001179; FKBP_PPIase.
 DR InterPro: IPR005215; Trig_fac.
 DR Pfam: PF00254; FKBP; 1
 DR TIGRfam: TIGR00115; tig; 1.
 DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.
 DR PROSITE: PS00454; FKBP_PPIASE_2; FALSE_NEG.
 DR PROSITE: PS50059; FKBP_PPIASE_3; 1.
 KW Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.
 FT DOMAIN 170 255
 FT SEQUENCE 444 AA: 50931 MW: 99704A2BA4E23F47 CRC64;
 SQ
 Query Match 12.9%; Score 72.5; DB 1; Length 444;
 Best Local Similarity 26.1%; Pred. No. 6.7;
 Matches 31; Conservative 18; Mismatches 31; Indels 39; Gaps 6;
 QY 10 PACIPTSSSKOLFSLFHYVOQPSGNGEKOYTTISHSSTLTQKCGQKMPVNTIIPTSQ 69
 DB 267 PEOFOSTSLKAFKSYFHKIME---NKQETIIQENN---QKIQFL-TWTKLIP--- 314
 QY 70 FPPASIKQITLPENKTLISLASPTOKNRI-----KENVT 104
 DB 315 FLPEALIK---LEANNRLKLQSOAEQYKIPFKELLSASNTITLLEODNRKIKAKENVT 370
 RESULT 10
 REF1_HUMAN
 ID REF1_HUMAN STANDARD; PRT; 979 AA.
 AC P22670;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE MHC class II regulatory factor REF1 (REF) (Enhancer factor C) (EF-C).
 GN REF1.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91071581; PubMed=2253877;

RA Relth W., Sanchez-Herrero C., Kober M., Silacci P., Berte C.,
RA Barras E., Mach B.;
RT "MHC class II regulatory factor RFX has a novel DNA-binding domain
RT and a functionally independent dimerization domain.";
RL Genes Dev. 4:1528-1540(1990).
RN [2]
RP IDENTITY BETWEEN RFX1 AND EF-C.
RX MEDLINE=94019311; PubMed=8413236;
RA Siegrist C.A., Durand B., Emery P., David E., Hearing P., Mach B.,
RA Relth W.;
RT "RFX1 is identical to enhancer factor C and functions as a
RT transactivator of the hepatitis B virus enhancer.";
RL Mol. Cell. Biol. 13:6375-6384(1993).
RN [3]
RP BINDING TO RPL30 PROMOTER.
RX MEDLINE=94040774; PubMed=8224874;
RA Safirany G., Perry R.P.;
RT "transcription factor RFX1 helps control the promoter of the mouse
RT ribosomal protein-encoding gene rpl30 by binding to its alpha
RT element.";
RL Gene 132:279-283(1993).
RN [4]
RP SHOWS THAT BLS IS NOT DUE TO RFX1.
RX MEDLINE=92375076; PubMed=1508204;
RA Sanchez-Herrero C., Relth W., Silacci P., Mach B.;
RT "The DNA-binding defect observed in major histocompatibility complex
RT class II regulatory mutants concerns only one member of a family of
RT complexes binding to the X boxes of class II promoters.";
RL Mol. Cell. Biol. 12:4076-4083(1992).
CC -1- FUNCTION: REGULATORY FACTOR ESSENTIAL FOR MHC CLASS II GENE
CC EXPRESSION. BINDS TO THE X BOXES OF MHC CLASS II GENES. ALSO BINDS
CC TO AN INVERTED REPEAT (ENH1) REQUIRED FOR HEPATITIS B VIRUS GENES
CC EXPRESSION AND TO THE MOST UPSTREAM ELEMENT (ALPHA) OF THE RPL30
CC PROMOTER.
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE RFX FAMILY.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58964; GAA41730.1; -;
DR EMBL: A20498; GAA41506.1; -;
DR PIR: A35913; A35913.
DR TRANSFAC: T00909; -;
DR TRANSFAC: T01673; -;
DR Genew: HGNC: 9982; RFX1.
DR MIM: 600006; -;
DR InterPro: IPR003150; RFX_DNA_binding.
DR Pfam: PF02257; RFX_DNA_binding; 1.
KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 381 411 GLY-RICH.
FT DNA_BIND 438 528 EXPERIMENTALLY DEDUCED.
FT DOMAIN 920 936 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 744 979 NECESSARY FOR DIMERIZATION.
FT SEQUENCE 979 AA; 104728 MW; 556151f886c6a9a2 CRC64;
SO
Query Match 12.7%; Score 71.5; DB 1; Length 979;
Best Local Similarity 22.8%; Pred. No. 21;
Matches 34; Conservative 17; Mismatches 43; Indels 55; Gaps 6;
QY 1 AOPGVUSOPAGIPT-----GSSSKOLFSLFHVVOQ 31
DB 80 AVPA--SOPGAPTPSPAPOQYIVTVSESGAMRASSTVSEASPGSTASQGVPTQVVOQ 137
QY 32 PSNGNEKYVTIIS-----HSTILIOKCGOKTAMPVNTIIPSPAPPSAIKQITLPGNK 85
DB 138 VQGVOQRLIVOTSVQAKPGHVSPIQL-----TNTQVPOQALPTQRLVVOQSAAPSGK 188

QY 86 --ILSLQA-----SPROKNRIKENVTS 105
DB 189 GGOVSLTVHGHOQVHVSPEQSPVQANSSS 217
RESULT 11
PODX_HUMAN STANDARD; PRT: 528 AA.
AC 000592;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Podocalyxin-like protein 1 precursor.
GN PODXL OR PCPL1 OR PCPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97332652; PubMed=9188463;
RA Kershaw D.B., Beck S.G., Wharram B.L., Wiggins J.E., Goyal M.,
RA Thomas P.E., Wiggins R.C.;
RT "Molecular cloning and characterization of human podocalyxin-like
RT protein. Orthologous relationship to rabbit PCPL1 and rat
RT podocalyxin.";
RL J. Biol. Chem. 272:15708-15714(1997).
CC -1- FUNCTION: Functions as an antiadhesin that maintains an open
CC filtration pathway between neighboring foot processes in the
CC podocyte by charge repulsion.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Podocyte).
CC -1- TISSUE SPECIFICITY: Glomerular epithelium cell (podocyte).
CC -1- PTM: Sialoglycoprotein.
CC -1- SIMILARITY: BELONGS TO THE PODCALYXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U97519; AAB61574.1; -;
DR Genew: HGNC: 9171; PODXL.
DR MIM: 602632; -;
KW Glycoprotein; Signal; Transmembrane; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 23 528
FT DOMAIN 23 431
FT TRANSMEM 432 452
FT DOMAIN 453 528
FT DOMAIN 22 304
FT CARBOHYD 35 35
FT CARBOHYD 45 45
FT CARBOHYD 106 106
FT CARBOHYD 146 146
FT CARBOHYD 330 330
FT VARIANT 62 62
FT VARIANT 62 62
FT VARIANT 196 196
FT SEQUENCE 528 AA; 55595 MW; 22A32B291F008D56 CRC64;
SO
Query Match 12.6%; Score 71; DB 1; Length 528;
Best Local Similarity 25.5%; Pred. No. 12;
Matches 27; Conservative 18; Mismatches 35; Indels 26; Gaps 5;
QY 5 PYLSOPAGIPTGSSSKOLFSLFHVVOQPSGNGEKV-----TTIHSSITLTIKCGOKT- 58
DB 194 PTLTHPVATPTSSGHD-----HLMKISSSSSTVAIPGYTFISPGMTTTLPSVSISORTQ 247

OY 59 -----MPVNTIITPSO-----PPASILKQITLPGNKILSLQASPT 94
 DB 248 QISSQMPASSTAPSSQETVQPISPATLRPTLPE-----TWSSSPT 289

RESULT 12
 NPT5_HUMAN STANDARD; PRT; 1531 AA.
 AC 094916; 095693; Q9UN18;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nuclear factor of activated T cells 5 (T cell transcription factor
 NFAT5) (NF-AT5) (tonicity-responsive enhancer-binding protein) (Tone-
 binding protein) (ToneBP).
 GN NFAT5 OR TONEBP OR KIA0827.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM C).
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
 Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:355-364(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC TISSUE=Brain;
 RX MEDLINE=99307389; PubMed=10377394;
 RA Lopez-Rodriguez C., Aramburu J., Rakeman A.S., Rao A.;
 RT "NFAT5, a constitutively nuclear NFAT protein that does not cooperate
 RT with Fos and Jun.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:7214-7219(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC TISSUE=Brain;
 RX MEDLINE=99162641; PubMed=10051678;
 RA Miyakawa H., Woo S.K., Dahl S.C., Handler J.S., Kwon H.M.;
 RT "Tonicity-responsive enhancer binding protein, a rel-like protein that
 RT stimulates transcription in response to hypertonicity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2538-2542(1999).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Brain;
 RA Lopez-Rodriguez C., Aramburu J., Rakeman A.S., Copeland N.G.,
 RT Gilbert D.J., Thomas S., Distche C., Jenkins N.A., Rao A.;
 RT "NFAT5: The NF-AT family of transcription factors expands in a new
 RT direction.";
 RL Cold Spring Harb. Symp. Quant. Biol. 64:517-526(1999).
 RN [5]
 RP SEQUENCE OF 675-1531 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20029268; PubMed=10565538;
 RA Zuenke C., Kiehl R., Johannsmeier A., Grzeschik K.H., Schwinger E.;
 RT "Isolation and characterization of novel CAG repeat containing genes
 RT expressed in human brain.";
 RL DNA Seq. 10:1-6(1999).
 CC -1- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF GENES.
 CC REGULATES HYPERTONICITY-INDUCED CELLULAR ACCUMULATION OF
 CC OSMOLYTES.
 CC -1- SUBUNIT: DOES NOT BIND WITH FOS AND JUN TRANSCRIPTION FACTORS. BUT
 CC MIGHT BE CAPABLE OF FORMING STABLE DIMERS WITH DNA ELEMENTS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B AND C (SHOWN HERE); MAY BE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE, BRAIN,
 CC HEART AND PERIPHERAL BLOOD LEUCOCYTES. ALSO EXPRESSED IN PLACENTA,
 CC LUNG, LIVER, KIDNEY, PANCREAS, SPLEEN, THYMUS, PROSTATE, TESTIS,
 CC OVARY, SMALL INTESTINE AND COLON.

CC -1- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.
 CC -1- CAUTION: REF.5 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1165
 CC ONWARD DUE TO A FRAMESHIFT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB020634; BA74850.1; -
 DR EMBL: AF089824; AAD18136.1; -
 DR EMBL: AF134870; AAD38360.1; -
 DR EMBL: 297016; CAB09693.1; ALT_FRAME.
 DR EMBL: AF163836; AAD48441.1; -
 DR HSSP: Q00653; 1A3Q.
 DR TRANSFAC: T04940; -
 DR GeneW: HGNC:7774; NFAT5.
 DR MIM: 604708; -
 DR InterPro: IPR002909; IPT_TIG.
 DR InterPro: IPR000451; NF_Rel_dor_fam.
 DR Pfam: PF01833; TIG: 1.
 DR SMART: SM00429; IPT_1.
 DR PROSITE: PS01204; REL_1; FALSE_NEG.
 DR PROSITE: PS0254; REL_2; 1.
 DR KW Transcription regulation; Activator; Nuclear protein; DNA-binding;
 KW Alternative splicing.
 FT DOMAIN 69 100 SER-RICH.
 FT DOMAIN 293 300 DNA-BINDING.
 FT DOMAIN 739 743 POLY-GLN.
 FT DOMAIN 879 888 POLY-GLN.
 FT DOMAIN 966 971 POLY-THR.
 FT DOMAIN 1248 1266 POLY-GLN.
 FT VARSPIC 1 76 MISSING (IN ISOFORM A).
 FT VARSPIC 1 47 MISSING (IN ISOFORM B).
 FT CONFLICT 1369 1369 E -> D (IN REF. 5).
 FT SEQUENCE 1531 AA; 165764 MW; A68C6808BDA6F69E CRC64;
 SQ
 Query Match 12.6%; Score 71; DB 1; Length 1531;
 Best Local Similarity 29.3%; Pred. No. 41;
 Matches 29; Conservative 18; Mismatches 28; Indels 24; Gaps 6;
 OY 1 AAGPGVLSQPA-----GIPTSSSKQL-FSLFH--VVOQPSGNGEKQVYTRISHS-STLTIG 52
 DB 975 ATRGTWTFORTSSGGDEGTQAKQIQNSVFGTMVQHQHSGDNQPVNLFSSSRKMSMSVQ 1034
 OY 53 KCGQKTPVNTIITPSQPPASILKQITLPGNKILSLQ 91
 DB 1035 NSG-----TQQDGNGLFQQ---GNEWMSLQS 1057

RESULT 13
 DP3X_HAEIN STANDARD; PRT; 688 AA.
 ID DP3X_HAEIN
 AC P43746;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA polymerase III subunit gamma/tau (EC 2.7.7.7).
 GN DNAX OR H11229.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,	
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,	
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,	
RA	Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,	
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,	
RA	Venter J.C.:	
RT	*Whole-genome random sequencing and assembly of Haemophilus Influenzae	
RT	Rd.":	
RL	Science 269:496-512(1995).	
CC	-1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME	
CC	RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.	
CC	THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.	
CC	-1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate	
CC	+ [DNA](N).	
CC	-1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA	
CC	CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE	
CC	DIMERIZATION TO FORM THE POLIIT' COMPLEX. POLIIT' ASSOCIATES WITH	
CC	THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,	
CC	AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
CC	EMBL; U32802; AAC22882.1; -	
DR	TIGR: H11229; -	
DR	InterPro: IPR003959; AAA_AtPase_cent.	
DR	InterPro: IPR000862; RFCdomain.	
DR	Pfam: PF00004; AAA; 1.	
KW	Transferase; DNA-directed DNA polymerase; DNA replication;	
KW	ATP-binding; Complete proteome.	
FT	NP_BIND 45 52 ATP (POTENTIAL).	
FT	SEQUENCE 688 AA; 77042 MW; DD3401A54C6C9A6F CRC64;	
CC	-----	
CC	Query Match 12.5%; Score 70; DB 1; Length 688;	
CC	Best Local Similarity 25.2%; Pred. No. 20;	
CC	Matches 31; Conservative 18; Mismatches 44; Indels 30; Gaps 5;	
QY	5 PVL5Q-----PAGIPGSSSKQLFSLFHVQVQSPSGNEKO-----VTT 42	
Db	: : : : : : : : : : : : :	
Db	390 PVL5QSIKSAYSQAKPNKTSIPNLASLSALDALEHLTQLEN--QERQEHKAESLAVVSET 447	
QY	43 ISHSSTLTIOCKGOKTPVNTIPTSQF--PPASILKQITLPGNKILSLQASTQKNRIK 100	
Db	448 LHHIQEULDEEKSHKM-----TALPVREMTPEKPKHIEKPTLP5NAAQAQPKNSTEENS 503	
QY	101 ENV 103	
Db	::	
Db	504 DNV 506	
Db	::	
RESULT 14		
SMA4_MOUSE		
ID	SN44_MOUSE STANDARD; PRT; 551 AA.	
AC	P57471; O9CW56;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Mothers against decapentaplegic homolog 4 (SMAD 4) (Mothers against	
DE	DPP homolog 4) (Deletion target in pancreatic carcinoma 4 homolog)	
DE	(Smad4).	
DE	MADH4 OR DPC4.	
GN	Mus musculus (Mouse).	
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OC	NCBI_TaxId=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=A/J; TISSUE=Lung;	


```
FT DOMAIN 320 529 DMB.
FT DOMAIN 450 465 POLY-ALA.
FT CONFLICT 257 257 S -> A (IN REF. 2).
FT CONFLICT 292 292 R -> P (IN REF. 2).
SQ SEQUENCE 551 AA; 60417 MW; 0835EF8D9C1C980 CRC64;

Query Match 12.3%; Score 69; DB 1; Length 551;
Best Local Similarity 35.9%; Pred. No. 19;
Matches 23; Conservative 6; Mismatches 31; Indels 4; Gaps 2;

QY 8 SQPAGIPTGSSSKQLFSLFHVQV-QPSGGNEKQVTTISHSTLTQKCGKQKMPVNTII 65
DB 222 SQPASTILASHSEGLLIQIASQPQGGQNGFTAQPSYHINSTTT--WTGSRATPTPNL 279

QY 66 PTSQ 69
DB 280 PHHQ 283

RESULT 15
RGSK_MOUSE STANDARD; PRT; 239 AA.
AC Q9QZB1; Q9QZB2; Q9CUV8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Regulator of G-protein signaling 20 (RGS20) (Regulator of G-protein
DE signaling 21).
GN RGS20 OR RGS21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ARG-32.
RC STRAIN=129/B6, and BALB/c;
RA Barker S.A., Wang J., Ross E.M.;
RA "A mouse ortholog of RGS21.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 5-239 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO
CC THEIR INACTIVE GDP-BOUND FORM. BINDS SELECTELY TO G(2)-ALPHA AND
CC IS INHIBITED BY THE PHOSPHORYLATION AND PALMITOYLATION OF THE G-
CC PROTEIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- PTM: FATTY ACYLATED. HEAVILY PALMITOYLATED IN THE CYSTEINE STRING
CC MOTIF (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 RGS DOMAIN.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF191554; AAF05757.1; -.
DR EMBL; AF191552; AAF05756.1; -.
DR EMBL; AK013773; BAB28987.1; -.
DR HSSP; P49795; 1CMZ.
DR MGD; MGI:1929866; Rgs20.
DR InterPro; IPR000342; Regl_Gprotein.
DR Pfam; PF00615; RGS; 1.
DR PRINTS; PR01301; RGS-PROTEIN.
DR ProDom; PD001580; Reg_of_prg; 1.
DR SMART; SM00313; RGS; 1.
DR PROSITE; PS0132; RGS; 1.
KW Signal transduction inhibitor; Membrane; Lipoprotein; Palmitate;
KW Polymorphism.
FT DOMAIN 59 71 POLY-CYS.
FT DOMAIN 113 229 RGS.
FT VARIANT 32 32 M -> R (IN BALB/C).
SQ SEQUENCE 239 AA; 26986 MW; F383923163A44D18 CRC64;

Query Match 12.2%; Score 68.5; DB 1; Length 239;
Best Local Similarity 21.1%; Pred. No. 8;
Matches 28; Conservative 23; Mismatches 53; Indels 29; Gaps 4;

QY 1 AQGPVLSQPAIGTSSSKQLFSLFHVQVQPSGGNEKQVTTISHS----- 46
DB 12 ASPSASPADP-GLPEGSEITEM-----RMKQCGSETQGPAPSQQGRGNACCCFWCC. 65

QY 47 -----STLTI-----QKCGKQKMPVNTIIISQPPASILKQITLPGNKILSLQASTQKN 97
DB 66 CCTCCLTVRNOEDQRPQRASHEIRTDIPACEESPTTLEEVCWAQSFNDLMVTPAGRN 125

QY 98 RIKENVTSFRDE 110
DB 126 AFREFLTFESEE 138

Search completed: February 16, 2003, 21:56:27
Job time : 6.11991 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:31:01 ; Search time 7.72591 Seconds
(without alignments)
2933.659 Million cell updates/sec

Title: US-09-763-909-2_COPY_443_552
Perfect score: 562
Sequence: 1 AQP6PVLSPAGIPTGSSK.....ASPTQKNRIKENTVSCFRDE 110

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvrius.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98.5	17.5	662	11	Q91WW6 mus musculus
2	84.5	15.0	624	4	Q96JG5
3	81.5	14.5	643	3	Q94233 kluyveromyc
4	80.5	14.3	410	3	Q9UTK3 schizosacch
5	79.5	14.1	3726	5	Q9VFL1 drosophila
6	78.5	14.0	506	5	Q8WQ86 dictyostell
7	78.5	14.0	612	5	Q8T0A8 drosophila
8	78.5	14.0	622	5	Q9W4F6 drosophila
9	78	13.9	346	16	Q9VRE1 delinococcu
10	77.5	13.8	545	2	Q52188 enterococcu
11	77.5	13.8	545	2	Q52193 enterococcu
12	77.5	13.8	751	5	Q960R2 drosophila
13	77	13.7	918	13	Q9DG67 gallus gall
14	76	13.5	403	3	Q42963 schizosacch
15	76	13.5	696	4	Q9H9R3 homo sapien
16	75.5	13.4	754	16	Q8XTG8 raistonias

17	74.5	13.3	2454	5	Q8T2G3 dictyostell
18	74	13.2	524	11	Q9EQY1 rattus norv
19	74	13.2	579	11	Q9EQY2 rattus norv
20	74	13.2	615	11	Q9EQY3 rattus norv
21	74	13.2	1844	5	Q22579 caenorhabdi
22	72.5	12.9	545	2	Q51643 enterococcu
23	72	12.8	327	4	Q9H683 homo sapien
24	72	12.8	370	11	Q9D3V7 mus musculu
25	72	12.8	1453	4	Q9Y6T1 homo sapien
26	72	12.8	1608	4	Q96RK0 homo sapien
27	72	12.8	2296	5	Q9VNE0 drosophila
28	71.5	12.7	554	6	Q95L89 bos taurus
29	71.5	12.7	580	6	Q8WML4 bos taurus
30	71.5	12.7	642	3	Q60129 schizosacch
31	71	12.6	783	12	Q91331 cercopithic
32	71	12.6	1484	4	Q96QH3 homo sapien
33	71	12.6	2766	11	Q9QZR8 rattus norv
34	70.5	12.5	157	16	Q8REV0 fusobacteri
35	70.5	12.5	397	10	Q94J09 arabidopsi
36	70.5	12.5	487	10	Q9SY1 arabidopsi
37	70.5	12.5	865	16	Q8YXD5 anabaena sp
38	70	12.5	218	10	Q43395 brassica na
39	70	12.5	410	4	Q95056 homo sapien
40	70	12.5	410	4	Q9UE14 homo sapien
41	70	12.5	424	12	Q99CX6 bovine herp
42	70	12.5	441	5	Q9XUS2 caenorhabdi
43	70	12.5	649	13	Q9DDC0 pleurodeles
44	70	12.5	1072	4	Q9Y4G7 homo sapien
45	70	12.5	1484	11	P70098 cricetus

ALIGNMENTS

RESULT 1

Q91WW6	PRELIMINARY;	PRT;	662 AA.
ID	Q91WW6		
AC	Q91WW6		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	TATA-binding protein associated factor TAFII135 (Fragment).		
GN	TAF4A		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BALB/C; TISSUE=PO BRAIN;		
RA	Metsis M., Brunkhorst A., Neuman T.;		
RT	*Cell Type Specific Expression of the TFIID Component TAFII135 in the		
RT	Nervous System.*;		
RL	Exp. Cell Res. 0:0-0(2001).		
DR	EMBL; AY038601; AAK94779.1;		
DR	MGD; MGI:2152346; Taf4a.		
FT	NON_TER		
SQ	SEQUENCE 662 AA; 71398 MW; 91A75F38CB0D0DA4 CRC64;		

Query Match 17.5% Score 98.5; DB 11; Length 662;
Best Local Similarity 30.6%; Pred. No. 0.03; Mismatches 43; Indels 25; Gaps 7;
Matches 36; Conservative 18;

Qy	2	QPGVLS--QPAGIPTGSSSKQLSFLHVVQPSGGN---EKQVTVISHSSTLTITKQCG 56
Db	293	QP-PVISLTQPTQGVG---KQAPPTPLVIQPPKPGALIRPPQVTLTQTPWALRQ--- 345
Qy	57	KTMVNTIIP-----SQFPASILKQIYLPNGKIL---SLQASPTQKNRIKENTVSC 106
Db	346	---PHNRIMLTTPQOIQLNQLPVPVVKPTVLPCTKALSTVSAQAAAAAKNKLKPGGGS 402
Qy	107	FRDE 110

Best Local Similarity 20.8%; Pred. No. 2.1;
Matches 27; Conservative 24; Mismatches 42; Indels 37; Gaps

QY 12 GIPTGSSKQLFSLFHVVOQPSGGNEK-----QVTTISHSSTLTQ-----KCGQKT 58
DB 142 GVPATSSSSSSSTIPLTNCKSTNAKSKAIOVQTTTITHTQPKPIAKIAPACKKQV 201

QY 59 MPVNTIIPTSQFPFASILKQITL-----PGNKIL---SIQASPT 94
DB 202 MPITLKPASAKMESAKVKPQVKVQEQDLVDLTHIQSQAPVEPKKPGSGTALKKEIKLKM 261

QY 95 QKNRIKENVT 104
DB 262 ENNKALKELT 271

RESULT 4
Q9UTK3
ID Q9UTK3 PRELIMINARY; PRT; 410 AA.
AC Q9UTK3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative nicotinate phosphoribosyltransferase.
GN SPAC1486.06.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133357; CAB62416.1; -
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 410 AA; 46694 MW; E4230ADD9ABA6077 CRC64;

Query Match 14.3%; Score 80.5; DB 3; Length 410;
Best Local Similarity 24.8%; Pred. No. 1.6;
Matches 29; Conservative 21; Mismatches 38; Indels 29; Gaps

QY 15 TGSSSKQLFSLFHVVOQPSG-----GNEKQVTTISHSSTLTQ----- 51
DB 273 TANSADDLANVFGVGRQDSCAEYIEKVVHYKHSIGVDSTKVIHSDALNVRDCLY 332

QY 52 ---QKCGQKT-MPVNTIIPTSQFPFASILKQITLPGNKILSL-QASPTQKNRIKENV 103
DB 333 KYCEKGIKSAFGIGTNL-TSDFQKSNPSEVSKPMNIVIKLFSAEKTKAVKISDDI 388

RESULT 5
Q9VFL1
ID Q9VFL1 PRELIMINARY; PRT; 3726 AA.
AC Q9VFL1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trx protein.
GN TRX OR CG8651.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner S., Henderson S.N.,

Best Local Similarity 15.0%; Score 84.5; DB 4; Length 624;
Matches 37; Conservative 11; Mismatches 41; Indels 39; Gaps

QY 3 PGPVLSOPAGIP-----TGSSSKQLFSLFHVVOQPSGGNEKQVTTISHSSTLTQKCGQK 57
DB 302 PGPASCQPGRCPLSPSATCDT-----RGVPQPSWGPQEAQAASSSPLEALEAC-UK 352

QY 58 TMPVNTIIPTSQFPFASILKQITLPGNKILSLQASPTQKNRIKEN----- 102
DB 353 GIPNGSSP-SQLPETS-CSQNPQPGD-----SRSKPELQPHRSHSEATREPVLP 403

QY 103 -VTSQFRD 109
DB 404 GLQSCVRD 411

RESULT 3
Q94233
ID Q94233 PRELIMINARY; PRT; 643 AA.
AC Q94233;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hap4p.
GN HAP4.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHS2359;
RX MEDLINE=99195821; PubMed=10096087;
RA Bourgaud D., Nguyen C.C., Bolotin-Fukuhara M.;
RA "HAP4, the glucose-repressed regulated subunit of the HAP
RT transcriptional complex involved in the fermentation-respiration
RT shift, has a functional homologue in the respiratory yeast
RT Kluyveromyces lactis.";
RL Mol. Microbiol. 31:1205-1215(1999).
DR EMBL; AF072675; AAD20134.1; -
SQ SEQUENCE 643 AA; 70841 MW; AFC8FDEC2FFCE3B7 CRC64;

Query Match 14.5%; Score 81.5; DB 3; Length 643;

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Piannkoch C., Baldwin D.,
 RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulo G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 DR EMBL: AE003704; AAF55041.2; -
 DR HSSP: P20393; 1A6Y.
 DR FlyBase: FBgn0003862; trx.
 DR InterPro: IPR003889; FYrich_C.
 DR InterPro: IPR003888; FYrich_N.
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR001965; Znf_PHD.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF00628; PHD; 3.
 DR Pfam: PF00856; SET; 1.
 DR SMART: SM00542; FYR1; 1.
 DR SMART: SM00541; FYR1; 1.
 DR SMART: SM00249; PHD; 4.
 DR SMART: SM00508; PostSET; 1.
 DR SMART: SM00184; RING; 3.
 DR SMART: SM00317; SET; 1.
 DR PROSITE: PS50280; SET; 1.
 SQ SEQUENCE 3726 AA; 40095 MW; E3DD8B8F062BD7796 CRC64;

Query Match 14.1%; Score 79.5; DB 5; Length 3726;
 Best Local Similarity 18.2%; Pred. No. 26;
 Matches 30; Conservative 24; Mismatches 50; Indels 61; Gaps 5;

QY 3 PGPVLSQAPAGITGSSSKQL--PSLFHVHVQPSGGNGKQV----- 40
 DB 2953 PKVTISQOR-IPATQTOQQQQAQMIHIPOQQPQQLOQQQVOVQSPMFIITLAEAPVQSQ 3011
 QY 41 -----TTTSHSTLTIOKCGTKGKTMVNTIIPTSQFPASILKQITLPGNKI 86
 DB 3012 FVMEPQALQEQELANRVQHFSTSSSSSSSSCSLFTNVNPMQQAQPSSTTSRTRNRV 3071
 QY 87 LSLQ-----ASPTQKNRIKE-----NVTSCF 107
 DB 3072 LPMQORQAPLNSNECPVSSPPPKPVEQPIIHQMTIASVSKY 3116

RESULT 6
 Q8WQ86 PRELIMINARY; PRT; 506 AA.
 AC Q8WQ86;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Microtubule-associated protein EBI.
 GN EBI.
 OS Dictyostellium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX2;
 RA Rehberg M., Graef R.;
 RT "deEBI is a permanent centrosomal resident in Dictyostellium and
 RT required for proper spindle formation.";
 RL Mol. Biol. Cell 12S:309a-309a(2001).
 DR EMBL: AJ426053; CAD19801.1; -
 DR InterPro: IPR001715; Calponin-like.
 DR InterPro: IPR004953; Calponin-like.
 DR Pfam: PF00307; CH; 1.
 DR Pfam: PF03271; EBI; 1.
 DR PROSITE: PS50021; CH; 1.
 SQ SEQUENCE 506 AA; 56983 MW; F20BC182BAAA639C CRC64;

Query Match 14.0%; Score 78.5; DB 5; Length 506;
 Best Local Similarity 32.0%; Pred. No. 3.3;
 Matches 32; Conservative 11; Mismatches 40; Indels 17; Gaps 4;

QY 1 AQPGP-VLSQAPAGITGSS-----SKOLFSLFHVHVQPSGGNGKQVTTTSHSTLTIOK 54
 DB 175 AKTPSSSVSKVPKPTSPSSISKVPKPTKPT-----PTTSTSTTTTSTP 228
 QY 55 GQKTMVNTIIPTSQFPASILKQITLPGNKILSLQASPT 94
 DB 229 PSTPKTNTPIPTGTGKPT-----LTQPTFKPTPKSVSPT 263

RESULT 7
 Q8TOA8 PRELIMINARY; PRT; 612 AA.
 AC Q8TOA8;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE LD15062p.
 GN CG4088.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Colniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY069439; AAL39584.1; -
 SQ SEQUENCE 612 AA; 65734 MW; E64C523CA2EA0623 CRC64;

Query Match 14.0%; Score 78.5; DB 5; Length 612;
 Best Local Similarity 27.5%; Pred. No. 4.1;
 Matches 28; Conservative 19; Mismatches 40; Indels 15; Gaps 5;

QY 5 PVLSPQAPAGITGSSSKQ--LPSLFHVHVQPSGGNGKQVTTTSHSTLTIOKCGTKMPVN 62
 DB 499 PLKINVLGGSGQGNKRAIFNSASSIOHENG-----VTTIVPASSLA---ASNQTAAMN 550

Qy 63 TIIP-TSQPPASILKQITLPGNKIISLQ-----ASPTQNRI 99
| | | : : : | : : | : : | : :
Db 551 AIARSTVITTPSSVGKKVTVPNPKFILLKPAKFVGPASANEV 592

RESULT 8
Q9W4F6

ID	Q9W4F6	PRELIMINARY;	PRT;	622 AA.
AC	Q9W4F6;			
DT	01-MAY-2000	(TREMBlrel. 13, Created)		
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)		
DT	01-MAY-2000	(TREMBlrel. 13, Last annotation update)		
DE	CG4068	protein.		
GN	CG4068.			
OS	Drosophila melanogaster	(Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Etheroptera; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
	[1]			
RN				
RP	SEQUENCE FROM N.A.			

RC	STRAIN=BERKELEY;	
RA	MEDLINE=20190606; PubMed=10731132;	
RA	Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.C., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Arrill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benson P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,	
RA	Burtis K.C., Busam C.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I.J., Dietz S.M.,	
RA	Donson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,	
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,	
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,	
RA	Shue B.C., Siden-Kiamos I., Sampson M., Skupski M.P., Smith T.,	
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,	
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:	
RT	"The genome sequence of <i>Drosophila melanogaster</i> ."	
RL	Science 287:2185-2195(2000).	
DR	EMBL: AE003432; AAF45997.1; -	
DR	Flybase: FBgn0029738; CG4068	
SQ	SEQUENCE 622 AA; 66817 MW; F6089D08FD8DB1A CRC64;	

Query Match	14.0%;	Score 78.5;	DB 5;	Length 622;
Best Local Similarity	27.5%;	Pred. NO. 4.2;		
Matches 28;	Conservative	19;	Mismatches 40;	Indels 15; Gaps 5;

QY 5 PVLSPACIGTSSSKQ--LFSLFHVVVQPSGGNEKQVTTISHSSTLTIQKCGOKTMPVN 62
|: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
Db 509 PLKINVGGSGQSNKMKRAIFNSASSIQHENG-----VTTIVPASSLA---ASNOTAAMN 560

QY 63 TIIP-TSQPPASILKQITLPGNKIISLQ-----ASPTQNRI 99
| | | : | : | : | : | : | : | :
DB 561 AIARSTVTTTSPSSVGKKVTVPNPKFILLKPAKFVGPASANEV 602

RESULT 9
O9RYE1

QRYEI	PRELIMINARY;	PRT;	346 AA.
AC	QRYEI;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Hypothetical protein DR0008.		
GN	DR0008.		
OS	Deinococcus radiodurans.		
OC	Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;		
OC	Deinococcaceae; Deinococcus.		
OX	NCBI_TaxID=1299;		
RN	{1}		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-RI;		
RX	MEDLINE=20036896; PubMed=10567266;		
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.J.		
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.J.,		
RA	Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.		
RA	Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,		
RA	Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.		
RA	Makotova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann		
RA	Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.		
RA	Fraser C.M.;		
RT	"Genome sequence of the radioresistant bacterium Deinococcus		
RT	radiodurans RI.";		
RL	Science 286:1571-1577(1999).		
RL	EMBL; AE001864; AAF09603.1; -.		
DR	TIGR; DR0008; -.		
KW	Hypothetical protein; Complete proteome.		
SQ	SEQUENCE 346 AA; 37095 MW; 1E348C3CC4D2BA8 CRC64;		

Query Match	13.9%	Score 78;	DB 16;	Length 346;
Best Local Similarity	22.0%	Pred. NO. 2.4;		
Matches 24;	Conservative 24;	Mismatches 45;	Indels 16;	Gaps 2;

[illegible]

RESULT 10
052188

Q52188	PRELIMINARY;	PRT;	545 AA.
AC	Q52188;		
DT	01-NOV-1996 (Tremblrel. 01, Created)		
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)		
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)		
DE	Trac precursor.		
DE	Trac.		
GN	Enterococcus faecalis (Streptococcus faecalis).		
OS	Enterococcus faecalis (Streptococcus faecalis).		
OG	Plasmid pPDI.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC	Enterococcaceae; Enterococcus.		
OX	NCBI_TaxID=1351;		
RN	[1]		

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE-9603293; PubMed-7559344;
RX Nakayama J., Yoshida K., Kobayashi H., Isogai A., Clewell D.,
RA Suzuki A.;
RT "Cloning and characterization of a region of enterococcus faecalis
RT plasmid pp1 encoding pheromone inhibitor (ipd), pheromone sensitivity
RT (trac), and pheromone shutdown (traB) genes";

```

RL J. Bacteriol. 177:5567-5573(1995).
DR EMBL: D28859; BAA06007.1; -.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 2.
DR PROSITE: PS01040; SBP_BACTERIAL_5; UNKNOWN_1.
KW Plasmid; Signal.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 545 AA; 60780 MW; 0338A2D15E091C73 CRC64;

Query Match 13.8%; Score 77.5; DB 2; Length 545;
Best Local Similarity 29.5%; Pred. No. 4.6;
Matches 26; Conservative 13; Mismatches 28; Indels 21; Gaps 4;

Qy 15 TGSSSKQLFSLFHVVOQPSGNEKQVTTISHSSTLTITKCGQKMPVNTIPTSQFPPAS 74
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 TASPVELFSAIKNAKETASG--KQV-----KDFLAVKSI G E K T L E I L V E P T Y F ---- 183

Qy 75 ILKQITLPGNKILSLQA-SPTQKNRIKE 101
| | | | | | | | | | | | | | | | | | | | | |
Db 184 -----TDLLSLTAYYPVQKAIKE 202

RESULT 11
Q52193 PRELIMINARY; PRT; 545 AA.
ID AC Q52193
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TRAC.
GN Enterococcus faecalis (Streptococcus faecalis).
OS Plasmid pPDI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032393; PubMed=7559344;
RA Nakayama J., Yoshida K., Kobayashi H., Isogai A., Clewell D.,
RA Suzuki A.;
RT "Cloning and characterization of a region of enterococcus faecalis
RT plasmid pPDI encoding pheromone inhibitor (ipd), pheromone sensitivity
RT (trac), and pheromone shutdown (trab) genes.";
RL J. Bacteriol. 177:5567-5573(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032394; PubMed=7559345;
RA Fujimoto S., Tomita H., Wakamatsu E., Tanimoto K., Ike Y.;
RT "Physical Mapping of the Conjugative Bacteriocin Plasmid pPDI of
RT Enterococcus faecalis and Identification of the Determinant Related to
RT the Pheromone Response.";
RL J. Bacteriol. 177:5574-5581(1995).
DR EMBL: D78016; BAA11195.1; -.
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 2.
DR PROSITE: PS01040; SBP_BACTERIAL_5; UNKNOWN_1.
KW Plasmid.
SQ SEQUENCE 545 AA; 60747 MW; 2277F1287C9D8291 CRC64;

Query Match 13.8%; Score 77.5; DB 2; Length 545;
Best Local Similarity 29.5%; Pred. No. 4.6;
Matches 26; Conservative 13; Mismatches 28; Indels 21; Gaps 4;

Qy 15 TGSSSKQLFSLFHVVOQPSGNEKQVTTISHSSTLTITKCGQKMPVNTIPTSQFPPAS 74
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 TASPVELFSAIKNAKETASG--KQV-----KDFLAVKSI G E K T L E I L V E P T Y F ---- 183

Qy 75 ILKQITLPGNKILSLQA-SPTQKNRIKE 101
| | | | | | | | | | | | | | | | | | | | | |
Db 184 -----TDLLSLTAYYPVQKAIKE 202

```

```

RESULT 12
Q960R2 PRELIMINARY; PRT; 751 AA.
ID AC Q960R2
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE LB39445p.
GN TRX OR CG8651.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051904; AAK93328.1; -.
DR FlyBase; FBgn0003862; trx.
DR InterPro: IPR001214; SET.
DR Pfam: PF00856; SET; 1.
DR PROSITE: PS0280; SET; 1.
SQ SEQUENCE 751 AA; 83669 MW; C9E299C65A623C57 CRC64;

Query Match 13.8%; Score 77.5; DB 5; Length 751;
Best Local Similarity 19.2%; Pred. No. 6.7;
Matches 20; Conservative 22; Mismatches 37; Indels 25; Gaps 3;

Qy 29 VQPSGGNEKQVTT-ISHSSTLTITKCGQKMPVNTIPTSQFPPASILKQITLPGNKIL 87
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 38 VMEPQALEQOELANRVQHFSTSSSSSSSSNCSLPTNVNPMQOQAPSTSSSTRPTNRVL 97

Qy 88 SLQ-----ASPTQKNRIKE-----NVTSCF 107
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 PMOQRPAPLNECPVVSPTPPKVEQPIHQWTSASVSKCY 141

RESULT 13
Q9DG67 PRELIMINARY; PRT; 918 AA.
ID AC Q9DG67
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Rad54b.
GN RAD54B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Yamaguchi-Iwai Y.;
RT "Rad54b, a Rad54-related gene product, is involved in recombinational
RL DNA repair.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF178529; AAG09308.1; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW ATP-binding; Helicase.

```

Search completed: February 16, 2003, 21:59:18
Job time : 12.7259 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:56:34 ; Search time 7.07238 Seconds
(without alignments)
1296.883 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359

Perfect score: 1773

Sequence: 1 GLTVKVPVSPAPRVSSGP.....KAGVVLHSGVPTAATGGTT 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2.6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2.6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2.6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2.6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2.6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2.6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2.6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2.6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2.6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2.6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2.6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2.6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2.6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2.6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	491	27.7	1023	9	US-09-893-519A-14
2	167.5	9.4	5179	9	US-10-025-380-1068
3	167.5	9.4	5179	10	US-09-922-217-1068
4	167.5	9.4	5179	10	US-09-833-263-1068
5	162	9.1	1367	10	US-09-801-368-108
6	152.5	8.6	386	10	US-09-864-761-35720
7	151	8.5	1169	10	US-09-801-368-106
8	149.5	8.4	2665	10	US-09-864-761-34248
9	144	8.1	941	12	US-10-124-557-14
10	144	8.1	1022	12	US-10-124-557-84
11	144	8.1	1038	12	US-10-124-557-74
12	144	8.1	1049	12	US-10-124-557-58
13	144	8.1	1140	12	US-10-124-557-104
14	144	8.1	1270	12	US-10-124-557-44
15	144	8.1	1311	12	US-10-124-557-42
16	144	8.1	1313	12	US-10-124-557-142
17	144	8.1	1314	12	US-10-124-557-50
18	144	8.1	1320	12	US-10-124-557-46
19	144	8.1	1320	12	US-10-124-557-60

20	144	8.1	1354	12	US-10-124-557-48
21	144	8.1	1361	12	US-10-124-557-40
22	144	8.1	1363	12	US-10-124-557-52
23	144	8.1	1404	12	US-10-124-557-2
24	144	8.1	1404	12	US-10-124-557-62
25	134	7.6	1075	10	US-09-801-368-110
26	129	7.3	1481	10	US-09-371-900-40
27	129	7.3	1481	10	US-09-924-417-60
28	127.5	7.2	821	10	US-09-841-132-195
29	127.5	7.2	1776	10	US-09-841-132-179
30	127	7.2	583	10	US-09-841-132-353
31	125	7.1	1322	10	US-09-801-368-114
32	124	7.0	511	10	US-09-864-761-34590
33	123.5	7.0	1601	10	US-09-862-027-40
34	121	6.8	1056	9	US-10-161-510-10
35	120	6.8	1848	9	US-09-839-996-6
36	118.5	6.7	1770	10	US-09-841-132-444
37	118	6.7	1537	10	US-09-801-368-104
38	118	6.7	2586	10	US-09-905-129-11
39	118	6.7	2586	10	US-09-905-129-14
40	118	6.7	2586	10	US-09-991-630-11
41	118	6.7	2586	10	US-09-991-630-11
42	118	6.7	2587	10	US-09-905-129-16
43	118	6.7	2587	10	US-09-991-630-16
44	118	6.7	2589	10	US-09-991-630-24
45	118	6.7	2597	10	US-09-905-129-2

ALIGNMENTS

RESULT 1
US-09-893-519A-14
; Sequence 14, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Tamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/IG548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
; DATABASE ENTRY DATE: 1997-06-25
; RELEVANT RESIDUES: (1)...(1023)

US-09-893-519A-14

Query Match 27.7%; Score 491; DB 9; Length 1023;
Best Local Similarity 38.4%; Pred. No. 6.5e-29;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTTIOPANLQPPGTVLKSNSSGGLM 61
DB 414 PTATSGIRATLTPTVLAPRLPOP-----PONPTNIQ---NFQLPPGMLVLRSENGQLL 464
QY 62 LVSPQOTVTR-----AETTSNITSRRPAPVNPANPQVTKTVPNSSSQLKKVAVTPVKKL 115
DB 465 MI-PQALAOQAOAHAPQOTMARPATPTSPAPVQISTVQAPGTPIIAR-QVTP-----518
QY 116 AIGTIVTVTVPKPSVQSVAVPTSVVTVPGKPLNTVT--TLKPSLSGASS-----TPS 168
DB 519 -----TTIIKQV---SQOATTVQPSATLQKSPGVQPOLVGGAAQTASLGTAFAVGTGPQ 571
QY 169 NE-PNLKAENSAVINLSPTMLENVKKCNFLAMLIKLACSGSOSPENGQWKKLVQOL 227
DB 572 RTVPGATTSSAATE-----TMENVKCKNFKLSTLIKLAASKQSTETAANVKELVQNL 625
QY 228 LDAKIEAEFTKRLKVELKSSPOPVLVFPKKSVALRQLLPNSOSFIQOCVQ-----TS 283
DB 626 LGKIEAEFTSLRYELNSSLSPQVLVFPKLSLALQRLPDLSAFLQSSQOQPPPPPTS 685
QY 284 SDMVIATCTTTVTSPVTVTVSSQSEKSIIVSGATAPRTVSV 327
DB 686 Q-----ATTALTAVLSSSVORTAGTKAATVTSALOPPVLSL 722

RESULT 2

US-10-025-380-1068
; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14

; CURRENT APPLICATION NUMBER: US/10/025.380

; CURRENT FILING DATE: 2001-12-19

; NUMBER OF SEQ ID NOS: 1129

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1068

; LENGTH: 5179

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-025-380-1068

Query Match 9.4%; Score 167.5; DB 9; Length 5179;
Best Local Similarity 22.08%; Pred. No. 0.00096;
Matches 89; Conservative 39; Mismatches 180; Indels 97; Gaps 14;

QY 2 TLVTKVAPVSAPPKVVSSGRLPAPQIVAVKAPNTTTIOPANLQPPGTVLKSNSSGGLM 61
DB 1607 TTTTTPPTTTPSPPTTPTIP-PTSTTLPLPTTTPSPPTTTPPTT-----TTPSPPT 1661

QY 62 LVSPQOTVTRAE-----TSNITSRRPAPVNPANPOT-----VKICTVNPSSSQL 103
DB 1662 TTTTSPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSTT-- 1719
QY 104 IKKAVTPVKKLAQIGTIVTVPKPSSVQSVAVP-----TSVVTVTPGKPLNTVTVLKPS 159
DB 1720 -----TPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTSPF 1773
QY 160 SLGASSTP-----SNEPNLKAENSAVAQIN--LSPTMLENVKCKNFLAMLIK 205
DB 1774 SYTTTTPCPLCNWGTGLDSCGPNFKPGDTELIGDVGCGWAANI-----1821
QY 206 LACSSQSP-----EMGNVKKLVQOLLDAKIEAEFTKRLKVELKSSPOPVLVFLKKS 260
DB 1822 -SCRATMYDPVPIGQLGQTVVCDVSVGLICKNEDQ-----KPG 1858
QY 261 VVALRQLLPNSOSFIQOCVQOTSSDMVIATCT--TTVTVSPVTVTVSSQSEKSIIVS 317
DB 1859 GVPMFAFLNYEINVOCCECVTQPTTMTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 1918
QY 318 GATAPRTVSVQTLNPLAGPVGAK-----AGVVTLHNSVGPATAAGGT 358
DB 1919 PTTTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPGT 1962

RESULT 3

US-09-922-217-1068

; Sequence 1068, Application US/09922217

; Patent No. US2002007641A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13

; CURRENT APPLICATION NUMBER: US/09/922,217

; CURRENT FILING DATE: 2001-08-03

; NUMBER OF SEQ ID NOS: 1124

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1068

; LENGTH: 5179

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-922-217-1068

Query Match 9.4%; Score 167.5; DB 10; Length 5179;
Best Local Similarity 22.0%; Pred. No. 0.00096;
Matches 89; Conservative 39; Mismatches 180; Indels 97; Gaps 14;

QY 2 TLVTKVAPVSAPPKVVSSGRLPAPQIVAVKAPNTTTIOPANLQPPGTVLKSNSSGGLM 61
DB 1607 TTTTTPPTTTPSPPTTPTIP-PTSTTLPLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTT 1661
QY 62 LVSPQOTVTRAE-----TSNITSRRPAPVNPANPOT-----VKICVNPSSSQL 103
DB 1662 TTTTSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSTT-- 1719
QY 104 IKKAVTPVKKLAQIGTIVTVPKPSSVQSVAVP-----TSVVTVTPGKPLNTVTVLKPS 159
DB 1720 -----TPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTSPF 1773
QY 160 SLGASSTP-----SNEPNLKAENSAVAQIN--LSPTMLENVKCKNFLAMLIK 205

;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

;; FILE REFERENCE: Aemica-X-1

;; CURRENT APPLICATION NUMBER: US/09/864,761

;; PRIOR FILING DATE: 2001-05-23

;; PRIOR APPLICATION NUMBER: US 60/180,312

;; PRIOR FILING DATE: 2000-02-04

;; PRIOR APPLICATION NUMBER: US 60/207,456

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: US 09/632,366

;; PRIOR FILING DATE: 2000-08-03

;; PRIOR APPLICATION NUMBER: GB 24263.6

;; PRIOR FILING DATE: 2000-10-04

;; PRIOR APPLICATION NUMBER: US 60/236,359

;; PRIOR FILING DATE: 2000-09-27

;; PRIOR APPLICATION NUMBER: PCT/US01/00666

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00667

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00664

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00669

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00665

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00668

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00663

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00662

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00661

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00670

;; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: US 60/234,687

;; PRIOR FILING DATE: 2000-09-21

;; PRIOR APPLICATION NUMBER: US 09/608,408

;; PRIOR FILING DATE: 2000-06-30

;; PRIOR APPLICATION NUMBER: US 09/774,203

;; PRIOR FILING DATE: 2001-01-29

;; NUMBER OF SEQ ID NOS: 49117

;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

;; SEQ ID NO 35720

;; LENGTH: 386

;; TYPE: PRT

;; ORGANISM: Homo sapiens

;; FEATURE:

;; OTHER INFORMATION: MAP TO AP000511.1

;; OTHER INFORMATION: EXPRESSED IN B474, SIGNAL = 1.7

;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3

;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7

;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4

;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8

;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1

;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5

;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5

;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7

;; OTHER INFORMATION: SWISSPROT HIT: P28968, EVALU 2.00e-05

US-09-864-761-35720

Query Match

Best Local Similarity 8.6%; Score 152.5; DB 10; Length 386;

Matches 85; Conservative 49; Mismatches 171; Indels 81; Gaps 13;

QY 2 TLTKVAPVAPPKVGSGRPLPAPQ-----IVAVKAPNTTIOFPANLQLPPTGLIK 54

DB 54 TTSSTVTPTTASTVSGSKNTACATSETTIAAITASEDTVS-----TQTSVIA 105

QY 55 SNSGP-----LMLVSPQOTVRAETTSNITSRPAPVAPNPQVKTCTVPNSSOLI- 104

DB 106 AESVPHATKTPTDITASVSATVPKNTPSVITSTPS-----TAPNTASKWT 154

RESULT 7

US-09-801-368-106

;; Sequence 106, Application US/09801368

;; Patent No. US20020128250A1

;; GENERAL INFORMATION:

;; APPLICANT: Busby, Robert

;; APPLICANT: Cali, Brian

;; APPLICANT: Hecht, Peter

;; APPLICANT: Holtzman, Doug

;; APPLICANT: Madden, Kevin

;; APPLICANT: Maxon, Mary

;; APPLICANT: Milne, Todd

;; APPLICANT: No. US20020128250A1man, Thea

;; APPLICANT: Royer, John

;; APPLICANT: Salama, Sofie

;; APPLICANT: Sherman, Amir

;; APPLICANT: Silva, Jeff

;; APPLICANT: Summers, Eric

;; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

;; FILE REFERENCE: 109272.147

;; CURRENT APPLICATION NUMBER: US/09/801,368

;; CURRENT FILING DATE: 2001-03-07

;; PRIOR APPLICATION NUMBER: US 09/487,558

;; PRIOR FILING DATE: 2000-01-19

;; PRIOR APPLICATION NUMBER: US 60/160,587

;; PRIOR FILING DATE: 1999-10-20

;; NUMBER OF SEQ ID NOS: 440

;; SOFTWARE: Patent in version 3.0

;; SEQ ID NO 106

;; LENGTH: 1169

;; TYPE: PRT

;; ORGANISM: Saccharomyces cerevisiae

US-09-801-368-106

Query Match

Best Local Similarity 21.2%; Score 151; DB 10; Length 1169;

Matches 87; Conservative 69; Mismatches 137; Indels 118; Gaps 18;

QY 4 VTKVAPVAPPKVGSGRPLPAPQIVAVKAPNTTIOF-PANLQLPPTGLIKNSQPLML 62

DB 689 VTSSVWSTPITTSSESASVTIL-----PSTITSEFKPSTMK-----TKVSISSSTNL 740

QY 63 VSPQOTVTRAETTSNITSRPAPV-----PANQTVKICTVNPSSSOLIKKAVTVP 112

DB 741 ITSYDTTSKDSITVSGSTSVSLISSISLPSYSASSEQIFHSSIVSSNGOALTSFSSTKV 800

QY 113 KK-----LAQIGTVTVTVPK-----PSSVQSVAVPTVVTVTPGK 149

DB 801 SSESSESHRTSPTTSSESGIKSGVEIESTTSFSFHEITSTASTSVQISQVTFPSP 860

QY 150 LNTVTLKPSSLG-----ASSTPSNPENLKAENSAVQINLSPTMLENVKKNFLAMLIK 205

Db 861 ISTVA---PRSTGLNSOTESTNSSKETMSSSENSASV-----893
Qy 206 LACSGSGSPENGQVKKLVLEOLLDAKEAE-----FTR---KLVELKSSQPHLVPLFK 258
Db 894 MPSSATSPKGT-----KVTSDTSSGFSRDRTTVYRMTSETPTN-----EQ 936
Qy 259 KSVVALRQLPNSOSFIQOCVQOQSSDMVIATCTTV-----TT-----SPVVTTVYSS 307
Db 937 TLLVSSCENSCS-----NTVSSAVSTATITINGITTEYTTWCPLSATELTITVSK 989
Qy 308 SOS-EKSIIVSGATAPRTVSVQTLNP--LAGPVGAKAGVVTLHSHVGTPTAAT 355
Db 990 LESEBKTTLITVTSCEGVCSETASPAIVSTATATATVNDVVTIVYSTWSPQAT 1040

RESULT 8
US-09-864-761-34248
; Sequence 34248, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034555.2

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EST_HUMAN HIT: AUI17052.1, EVALUO 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUO 3.00e-10
US-09-864-761-34248

Query Match 8.4%; Score 149.5; DB 10; Length 2665;
Best Local Similarity 23.7%; Pred. No. 0.0086;
Matches 85; Conservative 50; Mismatches 161; Indels 63; Gaps 14;

Qy 9 PVSAPPKVSSGPRLPAPQIVAVKAPNTTIOFPANLQLPPGTVLKSNKSGPLMLYS--PQ 66
Db 1841 PVTPP-----SDPSIPTPLPSVTA-----AKLSPVVASGGIPHOSPTKVTWITR 1887
Qy 67 QTVTRAETTSNITSRPAPVAPNPQTVKICTVPNSSSOLIKKAVATP--VKKLAQIGTIVVT 124
Db 1888 QEEPRASQTPS-----PALPPTDKASDVDTSSSTL-RKILMDPKYYSATSVTSTSVTT 1939
Qy 125 TVPKPSSVQSVAVPTSVVVTGKPLNTVTTIKPSSSLGASSTPSNEPNLKAENSAVQIN 184
Db 1940 AIAEPVSAAPCLHEAPPPVDSKKPLEKTA--PPTVNNSEIQASEVLVAADKERVAPV- 1996
Qy 185 LSPMLNENKCKNFAMLIKACSGSQSPENQGVNKKLVQQLDAAKEAEFTKRLVYE 244
Db 1997 IAPKITSVISR-----MPVSIDLENSOKITLAKPAPQTLTGLVSA-----2036
Qy 245 LKSSPQPHLVPLFKSVVALRQLPNSOSFIQOCVQOQSSDMVIATCTTTVTS- VVT 302
Db 2037 LTGLVNVSLP-----VNALGPVKGSVTTKLSLVSTPAGPVNVULGPVNVLTGPVNVLT 2091
Qy 303 TTVSSSQSEKSIIVSGATAPRTV--SVOTLNLPLAGPVGAKAGVVTLHSHVGTPTAATGGTT 359
Db 2092 TPVNAT-----VGTVNAAPGTVNAASAVNATASAVTGTAGAVTAASGGVTATTGTGT 2144

RESULT 9
US-10-124-557-14
; Sequence 14, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIORITY DATA:
; PRIORITY NUMBER: US 07/643,502

```
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 941 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-124-557-14

Query Match      8.1%; Score 144; DB 12; Length 941;
Best Local Similarity 20.3%; Pred. No. 0.0053;
Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 APVSAPPKVSSGPRLPAPQIVAVKAPNT-----TTIOFPANLQLPQGVTLI 53
DB 482 APTTPKEPAPTPKPEAPTPPKETAPTPKGAPATTLKEPAPTPPKKAPKELAPT--- 538
QY 54 KNSGFLMLVSPQQTIVRAETTSNITSRPAVPANPQTVKICTVPNSSSOLIKKVAVTPVK 113
DB 539 --TKETSTSDKPAPTPKGAPATTPKEPAPTPKPEAPTPKGATPTTLKEPAPTPK 596
QY 114 KLA-----QIGTVVTVTPKPSVQSVVA--VPTSVTVTPGKPLNTVTLKPSLSG 162
DB 597 KPAPKELAPTTTKGPTSTSDKPAPTPKETAPTPKPEAPTPPKKAPATTPETPPPTS 656
QY 163 ASSTP--SNEPNL--KAENSAAVQINLST--MLN-----VKKCKNFLAMLIKACSG 210
DB 657 EVSTPTTKETPTTIHKSPDSETPSELSAETPKALENSPKPEGVPTTKPAATKPEMTTA 716
QY 211 -----SOSPENGQNVKLVLEQLDAKIEAEFEKLYVELKSSQPQHLV 254
DB 717 KDKTTERDLRTPTETTTAAPKMTKETATTTKTESKITATT-----TQVSTTTQDPT 770
QY 255 PFLKKSVALRQLLNSQSFIOQCQQVQSSDMVIATCTTTTSPVTTTSSSQSEKSI 314
DB 771 PF-KITTLKTTTLAPK-----VTTTKKTIITTEIMNKPEETAKPKDRA 812
QY 315 IVSGATAPR 323
DB 813 TNSKATTPK 821

RESULT 10
US-10-124-557-84
; Sequence 84, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesnar, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
```

```
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Query Match      8.1%; Score 144; DB 12; Length 1022;
Best Local Similarity 20.3%; Pred. No. 0.006;
Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 APVSAPPKVSSGPRLPAPQIVAVKAPNT-----TTIOFPANLQLPQGVTLI 53
DB 563 APTTPKEPAPTPKPEAPTPPKETAPTPKGATPTTLKEPAPTPPKKAPKELAPT--- 619
QY 54 KNSGFLMLVSPQQTIVRAETTSNITSRPAVPANPQTVKICTVPNSSSOLIKKVAVTPVK 113
DB 620 --TKETSTSDKPAPTPKGAPATTPKEPAPTPKPEAPTPKGATPTTLKEPAPTPK 677
QY 114 KLA-----QIGTVVTVTPKPSVQSVVA--VPTSVTVTPGKPLNTVTLKPSLSG 162
DB 678 KPAPKELAPTTTKGPTSTSDKPAPTPKETAPTPKPEAPTPPKKAPATTPETPPPTS 737
QY 163 ASSTP--SNEPNL--KAENSAAVQINLST--MLN-----VKKCKNFLAMLIKACSG 210
DB 738 EVSTPTTKETPTTIHKSPDSETPSELSAETPKALENSPKPEGVPTTKPAATKPEMTTA 797
QY 211 -----SOSPENGQNVKLVLEQLDAKIEAEFEKLYVELKSSQPQHLV 254
DB 798 KDKTTERDLRTPTETTTAAPKMTKETATTTKTESKITATT-----TQVSTTTQDPT 851
QY 255 PFLKKSVALRQLLNSQSFIOQCQQVQSSDMVIATCTTTTSPVTTTSSSQSEKSI 314
DB 852 PF-KITTLKTTTLAPK-----VTTTKKTIITTEIMNKPEETAKPKDRA 893
QY 315 IVSGATAPR 323
DB 894 TNSKATTPK 902

RESULT 11
US-10-124-557-74
; Sequence 74, Application US/10124557
; Patent No. US20020137894A1
```

GENERAL INFORMATION:
 APPLICANT: Turner, Katherine
 Clark, Stephen C.
 Jacobs, Kenneth
 Hewick, Rodney M.
 Gesner, Thomas G.
 TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 NUMBER OF SEQUENCES: 143
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/124,557
 FILING DATE: 16-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cserr, Luann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 74:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1038 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 74:
 US-10-124-557-74

Query Match 8.1%; Score 144; DB 12; Length 1038;
 Best Local Similarity 20.3%; Pred. No. 0.0061;
 Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 APVSAPKVSGLPRLPAQIVAVKAPNT-----TTIQPANLQLPPGVLI 53
 DB 547 APTTPKEPATTPKEPATTPKGTAPTKGTAPTKLKEPATTPKPKAPKELAPTT--- 603

QY 54 KNSGGLMLVSPQQTIVTRAETTSNITSRPVAVPANPQVVKICTVFNSSQLIKKVAIVPVK 113
 DB 604 --TKEPTSTSDKPAITPKGTAPTKKEPATTPKKEPATTPKGTAPTKLKEPATTPK 661

QY 114 KLA-----QIGTVVTVTPKPSVQVA--VPTSVVTVTPGKPLNIVTVLKPSLSG 162
 DB 662 KPAPKELAPTTTKGPTSTSDKPAITPKGTAPTKKEPATTPKKEPATTPKPTPTTS 721

QY 163 ASSTP--SNEPNL--KAENSAVQINLSPT--MLEN-----VKCKCNFLAMLIKACSG 210
 DB 722 EVSPTTKETPTIHKSPDSESTSELSNEPTPKALENSPKPEPGVPTTKTAPTKPEMTTA 781

QY 211 -----SQSPMGONVKVLQELLDKAEAEFTKILYVELKSSQPQHLV 254
 DB 782 KDKTTERDLRTPPTTTAAAPKMKETATTEKTESKITATT-----TQVTSITTDIT 835

QY 255 PFLKSVVALRQLLPNSQSFIOQCQVQQTSSDMVIATCTTTTSPVVTITVSSQSEKSI 314
 DB 836 PF-KITTLKTTTLAPK-----VTTTKKTTITTEIMNKPEETAKPKDRA 877

QY 315 IVSGATAPR 323
 DB 878 TNSKATTPK 886

RESULT 12
 US-10-124-557-58
 ; Sequence 58, Application US/10124557
 ; Patent No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 Cambridgepark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 58:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1049 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
 US-10-124-557-58

Query Match 8.1%; Score 144; DB 12; Length 1049;
 Best Local Similarity 20.3%; Pred. No. 0.0062;
 Matches 75; Conservative 53; Mismatches 159; Indels 82; Gaps 12;

QY 8 APVSAPKVSGLPRLPAQIVAVKAPNT-----TTIQPANLQLPPGVLI 53
 DB 590 APTTPKEPATTPKEPATTPKGTAPTKGTAPTKLKEPATTPKPKAPKELAPTT--- 646

QY 54 KNSGGLMLVSPQQTIVTRAETTSNITSRPVAVPANPQVVKICTVFNSSQLIKKVAIVPVK 113
 DB 647 --TKEPTSTSDKPAITPKGTAPTKKEPATTPKKEPATTPKGTAPTKLKEPATTPK 704

Query Match 8.1%; Score 144; DB 12; Length 1140;
Best Local Similarity 20.3%; Pred. No. 0.0069;
Matches 75; Conservative 53; Mismatches 159; Indels 82

QY	211	-----SOSPENQONVKVLVOLLDAKIEAEFTRKLYVELKSSQPQHLV	254
DB	916	-----KDKITERDLRTPETTTAAKPMKETATTIEKTESKITATT-----	969
QY	255	PFLKKSVAALRQLLPNSQSFIOCCVQOQTSDDAVIATCTVTTPSVVTTVSSSQSKSI	314
DB	970	PF-KITTLKTTTLAPK-----VTTTKEKTTITTEIMNKPEETAKPDRA	1011

Db 1012 TNSKATTPK 1020

RESULT 14
US-10-124-557-44
; Sequence 44, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICATION NUMBER: 10124557
; PUBLICATION NUMBER: 20020137894A1
; INVENTOR: JAMES M. HARRIS
; ATTORNEY: JAMES M. HARRIS

STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

```

? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/10/124,557
? FILING DATE: 16-Apr-2002
? CLASSIFICATION: <unknown>
?
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/643,502
? FILING DATE: 18-JAN-1991
? APPLICATION NUMBER: US 07/546,114
? FILING DATE: 29-JUN-1990
? APPLICATION NUMBER: US 07/457,196
? FILING DATE: 29-DEC-1989
? APPLICATION NUMBER: US 07/390,901
? FILING DATE: 08-AUG-1989
? ATTORNEY/AGENT INFORMATION:
?
```


1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 22:01:05 ; Search time 126.688 Seconds
(without alignments)
1827.005 Million cell updates/sec

Title: US-09-763-909-2_COPY_1_359
Perfect score: 1773
Sequence: 1 GTLTKVAPVAPPKVSPP.....KAGVTLHSVGPTRATGGTT 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/PCTUS.COMB.pcp.*
- 2: /cgn2_6/ptodata/2/paa/US06.COMB.pcp.*
- 3: /cgn2_6/ptodata/2/paa/US07.COMB.pcp.*
- 4: /cgn2_6/ptodata/2/paa/US08.COMB.pcp.*
- 5: /cgn2_6/ptodata/2/paa/US081.COMB.pcp.*
- 6: /cgn2_6/ptodata/2/paa/US082.COMB.pcp.*
- 7: /cgn2_6/ptodata/2/paa/US083.COMB.pcp.*
- 8: /cgn2_6/ptodata/2/paa/US084.COMB.pcp.*
- 9: /cgn2_6/ptodata/2/paa/US085.COMB.pcp.*
- 10: /cgn2_6/ptodata/2/paa/US086.COMB.pcp.*
- 11: /cgn2_6/ptodata/2/paa/US087.COMB.pcp.*
- 12: /cgn2_6/ptodata/2/paa/US088.COMB.pcp.*
- 13: /cgn2_6/ptodata/2/paa/US089.COMB.pcp.*
- 14: /cgn2_6/ptodata/2/paa/US090.COMB.pcp.*
- 15: /cgn2_6/ptodata/2/paa/US091.COMB.pcp.*
- 16: /cgn2_6/ptodata/2/paa/US092.COMB.pcp.*
- 17: /cgn2_6/ptodata/2/paa/US093.COMB.pcp.*
- 18: /cgn2_6/ptodata/2/paa/US094.COMB.pcp.*
- 19: /cgn2_6/ptodata/2/paa/US095.COMB.pcp.*
- 20: /cgn2_6/ptodata/2/paa/US096.COMB.pcp.*
- 21: /cgn2_6/ptodata/2/paa/US097.COMB.pcp.*
- 22: /cgn2_6/ptodata/2/paa/US098.COMB.pcp.*
- 23: /cgn2_6/ptodata/2/paa/US099.COMB.pcp.*
- 24: /cgn2_6/ptodata/2/paa/US100.COMB.pcp.*
- 25: /cgn2_6/ptodata/2/paa/US101.COMB.pcp.*
- 26: /cgn2_6/ptodata/2/paa/US102.COMB.pcp.*
- 27: /cgn2_6/ptodata/2/paa/US60.COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|------------------------|-------------------|
| 1 | 1773 | 100.0 | 801 | 1 PCT-US01-08631-39827 | Sequence 39827, A |
| 2 | 1773 | 100.0 | 843 | 1 PCT-US02-29964-410 | Sequence 410, App |
| 3 | 1773 | 100.0 | 852 | 21 US-09-763-909-2 | Sequence 2, Appl1 |
| 4 | 1761.5 | 99.4 | 865 | 1 PCT-US02-25829-25 | Sequence 25, Appl |
| 5 | 1575 | 88.8 | 685 | 27 US-60-243-468-1271 | Sequence 1271, Ap |
| 6 | 491 | 27.7 | 737 | 1 PCT-US94-01114-16 | Sequence 16, Appl |

| | | | | | |
|----|-------|------|------|-------------------------|-------------------|
| 7 | 491 | 27.7 | 1023 | 1 PCT-US01-20592-14 | Sequence 14, Appl |
| 8 | 491 | 27.7 | 1023 | 1 PCT-US01-20592A-14 | Sequence 14, Appl |
| 9 | 491 | 27.7 | 1023 | 22 US-09-893-519A-14 | Sequence 14, Appl |
| 10 | 491 | 27.7 | 1051 | 1 PCT-US02-30474-1660 | Sequence 1660, Ap |
| 11 | 491 | 27.7 | 1051 | 27 US-60-324-631-1665 | Sequence 1665, Ap |
| 12 | 479 | 27.0 | 1083 | 1 PCT-US02-30474-3271 | Sequence 3271, Ap |
| 13 | 218 | 12.3 | 899 | 27 US-60-167-217-11401 | Sequence 11401, A |
| 14 | 218 | 12.3 | 921 | 27 US-60-173-464-9223 | Sequence 9223, Ap |
| 15 | 218 | 12.3 | 921 | 27 US-60-173-464-20611 | Sequence 20611, A |
| 16 | 216 | 12.2 | 921 | 1 PCT-US94-01114-2 | Sequence 2, Appl1 |
| 17 | 216 | 12.2 | 921 | 4 US-08-013-412-2 | Sequence 2, Appl1 |
| 18 | 216 | 12.2 | 921 | 20 US-09-614-150-11376 | Sequence 11376, A |
| 19 | 216 | 12.2 | 921 | 20 US-09-614-150-24957 | Sequence 24957, A |
| 20 | 216 | 12.2 | 921 | 27 US-60-191-637-11408 | Sequence 11408, A |
| 21 | 216 | 12.2 | 921 | 27 US-60-191-637-25058 | Sequence 25058, A |
| 22 | 216 | 12.2 | 921 | 27 US-60-191-681-8933 | Sequence 8933, Ap |
| 23 | 216 | 12.2 | 921 | 27 US-60-191-681-19739 | Sequence 19739, A |
| 24 | 182 | 10.3 | 1795 | 20 US-09-614-150-36210 | Sequence 36210, A |
| 25 | 182 | 10.3 | 1795 | 20 US-09-614-150-35824 | Sequence 35824, A |
| 26 | 182 | 10.3 | 1795 | 27 US-60-191-681-28036 | Sequence 28036, A |
| 27 | 175.5 | 9.9 | 708 | 1 PCT-US01-04098A-3624 | Sequence 3624, Ap |
| 28 | 175.5 | 9.9 | 709 | 1 PCT-US01-04098A-1656 | Sequence 1656, Ap |
| 29 | 175.5 | 9.9 | 709 | 20 US-09-629-469A-11876 | Sequence 11876, A |
| 30 | 175.5 | 9.9 | 709 | 20 US-09-629-469A-14832 | Sequence 14832, A |
| 31 | 175.5 | 9.9 | 881 | 20 US-09-611-526-3857 | Sequence 3857, Ap |
| 32 | 175.5 | 9.9 | 1294 | 1 PCT-US02-09921-1119 | Sequence 1119, Ap |
| 33 | 175.5 | 9.9 | 528 | 22 US-09-840-746-20 | Sequence 20, Appl |
| 34 | 173.5 | 9.8 | 528 | 22 US-60-185-361-566 | Sequence 566, App |
| 35 | 168 | 9.5 | 1953 | 27 US-60-185-361-494 | Sequence 494, App |
| 36 | 168 | 9.5 | 2947 | 27 US-60-185-361-938 | Sequence 938, App |
| 37 | 168 | 9.5 | 2947 | 20 US-09-641-377-697 | Sequence 697, App |
| 38 | 167.5 | 9.4 | 692 | 20 US-09-641-377-686 | Sequence 686, App |
| 39 | 167.5 | 9.4 | 717 | 20 US-10-221-279-12349 | Sequence 12349, A |
| 40 | 167.5 | 9.4 | 4742 | 26 US-02-17382-218 | Sequence 218, App |
| 41 | 167.5 | 9.4 | 5179 | 19 US-09-538-092-1258 | Sequence 1258, Ap |
| 42 | 167.5 | 9.4 | 5179 | 20 US-09-538-092-1068 | Sequence 1068, Ap |
| 43 | 167.5 | 9.4 | 5179 | 20 US-09-538-092-1068 | Sequence 1068, Ap |
| 44 | 167.5 | 9.4 | 5179 | 22 US-09-833-263-1068 | Sequence 1068, Ap |
| 45 | 167.5 | 9.4 | 5179 | 22 US-09-833-263-1068 | Sequence 1068, Ap |

ALIGNMENTS

RESULT 1
PCT-US01-08631-39827
; Sequence 39827, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 39827
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-39827

| | | | | |
|-----------------------|-----------------|---------------------|------------------|---------------------------|
| Query Match | 100.0% | Score 1773; | DB 1; | Length 801; |
| Best Local Similarity | 100.0%; | Pred. No. 4.6e-139; | | |
| Matches 359; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| QY | 1 | GTLTKVAPVSAPPKVSPP | PAPQIVAVKAPNTTTT | IQFPANLQLPPTGVLIKSNGPL 60 |
| Db | 1 | GTLTKVAPVSAPPKVSPP | PAPQIVAVKAPNTTTT | IQFPANLQLPPTGVLIKSNGPL 60 |

```
QY 61 MLVSPQQTVAETTSNITSRPAVPANPOTVKICTVPNSSQLIKKVAATPVKKLAQIGT 120
Db 61 MLVSPQQTVAETTSNITSRPAVPANPOTVKICTVPNSSQLIKKVAATPVKKLAQIGT 120
QY 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLGLASSTPSNEPNLKAENSA 180
Db 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLGLASSTPSNEPNLKAENSA 180
QY 181 VOINLSPMTLENVKKCNFLAMLILKACSGSPENGMQNVKKLVEQLDLDAKIEAEFTTRK 240
Db 181 VOINLSPMTLENVKKCNFLAMLILKACSGSPENGMQNVKKLVEQLDLDAKIEAEFTTRK 240
QY 241 LYVELKSSPOPHLVFELKKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTTSPV 300
Db 241 LYVELKSSPOPHLVFELKKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTTSPV 300
QY 301 VTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTT 359
Db 301 VTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTT 359
```

RESULT 2

```
PCT-US02-29964-410
; Sequence 410, Application PC/TUS0229964
; GENERAL INFORMATION:
```

```
; APPLICANT: Tang, Y. Tom
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aldong J.
; APPLICANT: Wang, Dunrui
; APPLICANT: Ghosh, Malabika
; APPLICANT: Asundi, Vinod
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle W.
; APPLICANT: Weng, Gezhi
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drmanac, Radolje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 809ACIP PCT
; CURRENT APPLICATION NUMBER: PCT/US02/29964
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 60/323,739
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 992
; SOFTWARE: pc_flg-genes version 6.0
; SEQ ID NO 410
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-29964-410
```

```
Query Match 100.0%; Score 1773; DB 1; Length 843;
Best Local Similarity 100.0%; Pred. No. 5e-139;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GTLVTKVAPVSAPPKVSQSGPRLPAPQIVAVKAPNTTTTIOFPANLQLPPGTVLIKSNGL 60
Db 43 GTLVTKVAPVSAPPKVSQSGPRLPAPQIVAVKAPNTTTTIOFPANLQLPPGTVLIKSNGL 102
QY 61 MLVSPQQTVAETTSNITSRPAVPANPOTVKICTVPNSSQLIKKVAATPVKKLAQIGT 120
Db 103 MLVSPQQTVAETTSNITSRPAVPANPOTVKICTVPNSSQLIKKVAATPVKKLAQIGT 162
QY 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLGLASSTPSNEPNLKAENSA 180
Db 163 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLGLASSTPSNEPNLKAENSA 222
QY 181 VOINLSPMTLENVKKCNFLAMLILKACSGSPENGMQNVKKLVEQLDLDAKIEAEFTTRK 240
Db 223 VOINLSPMTLENVKKCNFLAMLILKACSGSPENGMQNVKKLVEQLDLDAKIEAEFTTRK 282
QY 241 LYVELKSSPOPHLVFELKKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTTSPV 300
Db 283 LYVELKSSPOPHLVFELKKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTTSPV 342
QY 301 VTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTT 359
Db 343 VTTTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTT 401
```

RESULT 3

```
US-09-763-909-2
; Sequence 2, Application US/09763909
; GENERAL INFORMATION:
```

```
; APPLICANT: Dikstein, Rivka
; APPLICANT: Yamit-Hezi, Ayala
; TITLE OF INVENTION: A TRANSCRIPTION FACTOR TFIIID SUBUNIT,
; TITLE OF INVENTION: TAFII105, POLYPEPTIDES, DNA ENCODING THEREFOR AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS
; FILE REFERENCE: 13005/002001
; CURRENT APPLICATION NUMBER: US/09/763,909
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-763-909-2
```

```
Query Match 100.0%; Score 1773; DB 21; Length 852;
Best Local Similarity 100.0%; Pred. No. 5e-139;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GTLVTKVAPVSAPPKVSQSGPRLPAPQIVAVKAPNTTTTIOFPANLQLPPGTVLIKSNGL 60
Db 1 GTLVTKVAPVSAPPKVSQSGPRLPAPQIVAVKAPNTTTTIOFPANLQLPPGTVLIKSNGL 60
QY 61 MLVSPQQTVAETTSNITSRPAVPANPOTVKICTVPNSSQLIKKVAATPVKKLAQIGT 120
Db 61 MLVSPQQTVAETTSNITSRPAVPANPOTVKICTVPNSSQLIKKVAATPVKKLAQIGT 120
QY 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLGLASSTPSNEPNLKAENSA 180
Db 121 TVTIVTPKPSVQSVAVPTSVVTVTPGKPLNTVTLKPSLGLASSTPSNEPNLKAENSA 180
QY 181 VOINLSPMTLENVKKCNFLAMLILKACSGSPENGMQNVKKLVEQLDLDAKIEAEFTTRK 240
Db 181 VOINLSPMTLENVKKCNFLAMLILKACSGSPENGMQNVKKLVEQLDLDAKIEAEFTTRK 240
QY 241 LYVELKSSPOPHLVFELKKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTTSPV 300
Db 241 LYVELKSSPOPHLVFELKKSVVALRQLLPNSQSFIOQCVOQTSSDMVIATCTTTTSPV 300
```


QY 181 VOINLSPTMLENVKCKNFAMLIKACSGSPENQGNKKLVBLQDLDAKTEAEFEFTRK 240
DB 211 VOINLSPTMLENVKCKNFAMLIKACSGSPENQGNKKLVBLQDLDAKTEAEFEFTRK 270
QY 241 LYVELKSSPQPHLVFLPKKSVVALRQLLNSQSFQOCVQOQSSDMVIATCTTTTTSVP 300
DB 271 LYVELKSSPQPHLVFLPKKSVVALRQLLNSQSFQOCVQOQSSDMVIATCTTTTTSVP 330
QY 301 VTTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTLHSGVGTAAATGGTT 359
DB 331 VTTTSSSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTLHSGVGTAAATGGTT 389

RESULT 6
PCT-US94-01114-16
; Sequence 16, Application PC/TUS9401114
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAPS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESS: FLHR, HOUBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01114
; FILING DATE: 28-JAN-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-57650-2/AUT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01114-16

Query Match 27.7%; Score 491; DB 1; Length 737;
Best Local Similarity 38.4%; Pred. No. 1.4e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
QY 13 PKKVSSG-----PRLPAPQIVAVKAPNTTITQFPANLQLPFGVILKSNSSGGLM 61
DB 68 PTATSGIRATLPTVLAPRLPQ-----PQNPTNIQ---NFQLPPGMVLVRSNGQLL 118
QY 62 LVSPQOQTVTR-----AETTSNITSRPVAVPANTQVTKICTVPSNSSLKIKKVAVTPVKKL 115
DB 119 MI-PQALAQMOAQAHAQPTTMAPRPATPTISAPPVQISTVQAPCTPIIAR-QVTP---- 172

QY 116 AOIGTTVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVT--TLKPSSLGASS-----TPS 168
DB 173 ----TTIIKQV---SQAQTTVQPSATLQRSQVQPOLVGLGAAQASLGTATAVQTGTPO 225
QY 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFAMLIKACSGSPENQGNKKLVBLQDL 227
DB 226 RTVPGATTTSSAATE-----TMENYKCKNFELSTILKASSGKQSTETAANYKELVQNL 279
QY 228 LDKATEAEFEFTRKLYVELKSSPQPHLVFLPKKSVVALRQLLNSQSFQOCVQOQSSDM 283
DB 280 LDKATEAEFEFTRKLYVELKSSPQPHLVFLPKKSVVALRQLLNSQSFQOCVQOQSSDM 339
QY 284 SDMWIATCTTTTTSVPVTTTSSQSEKSIIVSGATAPRTVS 327
DB 340 Q-----ATTALTAVLSSSVQRTAGKTAATVTSALQPPVLSL 376

RESULT 7
PCT-US01-20592-14
; Sequence 14, Application PC/TUS0120592
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US1
; CURRENT APPLICATION NUMBER: PCT/US01/20592
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: human genbank accession #: CAA72189
; NAME/KEY: misc.feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
PCT-US01-20592-14

Query Match 27.7%; Score 491; DB 1; Length 1023;
Best Local Similarity 38.4%; Pred. No. 2.3e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;
QY 13 PKKVSSG-----PRLPAPQIVAVKAPNTTITQFPANLQLPFGVILKSNSSGGLM 61
DB 414 PTATSGIRATLPTVLAPRLPQ-----PQNPTNIQ---NFQLPPGMVLVRSNGQLL 464
QY 62 LVSPQOQTVTR-----AETTSNITSRPVAVPANTQVTKICTVPSNSSLKIKKVAVTPVKKL 115
DB 465 MI-PQALAQMOAQAHAQPTTMAPRPATPTISAPPVQISTVQAPCTPIIAR-QVTP---- 518
QY 116 AOIGTTVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVT--TLKPSSLGASS-----TPS 168
DB 519 ----TTIIKQV---SQAQTTVQPSATLQRSQVQPOLVGLGAAQASLGTATAVQTGTPO 571

QY 169 NE-PNLKAENSAAVQINLSPMTLENVKKCNFLAMLIKACSGSQSPMGONVKKLYEQL 227
DB 572 RTVPGATTSSAAE-----TMENVKKCNFLSTLIKASSGKQSTETAAANVKELVQNL 625
QY 228 LDAKIEAEFTSRKLYVELKSSPQPHLPFLKKSVALRQLLPNSQSFIOQCQVQ-----TS 283
DB 626 LDGKIEAEFTSRKLYVELKSSPQPHLPFLKKSVALRQLLPNSQSFIOQCQVQ-----TS 283
QY 284 SDMYIATCTTTVTSPVVTSSQSEKSIIVSGATAPRTVS 327
DB 686 Q-----ATTALTAVALSSSVQRTAGTAAATVTSALQPPVLSL 722

RESULT 8

PCT-US01-20592A-14

; Sequence 14, Application PC/TUS0120592A

; GENERAL INFORMATION:

; APPLICANT: ANADYS PHARMACEUTICALS, INC.

; APPLICANT: THOMPSON, Craig

; APPLICANT: MOORE, Jeffrey

; APPLICANT: BURMAN, Ed T.

; APPLICANT: BRADLEY, John

; APPLICANT: DESILVA, Thamara

; APPLICANT: HARRIS, Sandra

; APPLICANT: KOMARNITSKY, Svetlana

; APPLICANT: MENDILLO, Marc

; APPLICANT: MOORE, Daniel

; APPLICANT: MCCOY, Melissa

; APPLICANT: SANDERSON, Karen

; APPLICANT: HAQ, Tariq

; APPLICANT: ZHU, Shuhao

; APPLICANT: LONG, Fan

; APPLICANT: DAVIDOV, Eugene

; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE

; FILE REFERENCE: 0342/1G548-US1

; CURRENT APPLICATION NUMBER: PCT/US01/20592A

; CURRENT FILING DATE: 2001-06-28

; PRIOR APPLICATION NUMBER: US 60/215,164

; PRIOR FILING DATE: 2000-06-29

; PRIOR APPLICATION NUMBER: US 60/224,457

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 146

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 14

; LENGTH: 1023

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Corresponds to SEQ ID NO: 87

PCT-US01-20592A-14

Query Match 27.7%; Score 491; DB 1; Length 1023;
Best Local Similarity 38.4%; Pred. No. 2.3e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PKVSSG-----PRLPAQIVAVKAPNTTTTQFPANLQLPPGTVLKNSGGLM 61
DB 414 PTATTSGIRATLTPTVLAPRLPQ-----PQNPNTIQ---NFQLPGMWLVRSNGQLL 464
QY 62 LVSPQQTVTR-----AETTSNITSRAVPANPOTVKICTVPSNSSLIKKVAVTPVKKL 115
DB 465 MI-POQALAOQAOHAQOQPTMAPRATPTSAPPVQISTVQAPGTPIIAR-QVTP----- 518
QY 116 AQIGTTVTVTVPKSSVQSVAVPTSVVTPGKPLNTVT---TLKPSSILGASS-----TPS 168
DB 519 -----TTIIKQV---SQAQTTVPQSATLQSPGVQPOLVGLGAAQTASLGTATAVQTGTPQ 571
QY 169 NE-PNLKAENSAAVQINLSPMTLENVKKCNFLAMLIKACSGSQSPMGONVKKLYEQL 227
DB 572 RTVPGATTSSAAE-----TMENVKKCNFLSTLIKASSGKQSTETAAANVKELVQNL 625

QY 228 LDAKIEAEFTSRKLYVELKSSPQPHLPFLKKSVALRQLLPNSQSFIOQCQVQ-----TS 283
DB 626 LDGKIEAEFTSRKLYVELKSSPQPHLPFLKKSVALRQLLPNSQSFIOQCQVQ-----TS 283
QY 284 SDMYIATCTTTVTSPVVTSSQSEKSIIVSGATAPRTVS 327
DB 686 Q-----ATTALTAVALSSSVQRTAGTAAATVTSALQPPVLSL 722

RESULT 9

US-09-893-519A-14

; Sequence 14, Application US/09893519A

; GENERAL INFORMATION:

; APPLICANT: ANADYS PHARMACEUTICALS, INC.

; APPLICANT: THOMPSON, Craig

; APPLICANT: MOORE, Jeffrey

; APPLICANT: BURMAN, Ed T.

; APPLICANT: BRADLEY, John

; APPLICANT: DESILVA, Thamara

; APPLICANT: HARRIS, Sandra

; APPLICANT: KOMARNITSKY, Svetlana

; APPLICANT: MENDILLO, Marc

; APPLICANT: MOORE, Daniel

; APPLICANT: MCCOY, Melissa

; APPLICANT: SANDERSON, Karen

; APPLICANT: HAQ, Tariq

; APPLICANT: ZHU, Shuhao

; APPLICANT: LONG, Fan

; APPLICANT: DAVIDOV, Eugene

; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE

; FILE REFERENCE: 0342/1G548-US2

; CURRENT APPLICATION NUMBER: US/09/893,519A

; CURRENT FILING DATE: 2001-06-28

; PRIOR APPLICATION NUMBER: US 60/215,164

; PRIOR FILING DATE: 2000-06-29

; PRIOR APPLICATION NUMBER: US 60/224,457

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 146

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 14

; LENGTH: 1023

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Corresponds to SEQ ID NO: 87

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: Human Genbank/CAA72189

; DATABASE ENTRY DATE: 1997-06-25

; RELEVANT RESIDUES: (1)..(1023)

US-09-893-519A-14

Query Match 27.7%; Score 491; DB 22; Length 1023;
Best Local Similarity 38.4%; Pred. No. 2.3e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PKVSSG-----PRLPAQIVAVKAPNTTTTQFPANLQLPPGTVLKNSGGLM 61
DB 414 PTATTSGIRATLTPTVLAPRLPQ-----PQNPNTIQ---NFQLPGMWLVRSNGQLL 464
QY 62 LVSPQQTVTR-----AETTSNITSRAVPANPOTVKICTVPSNSSLIKKVAVTPVKKL 115
DB 465 MI-POQALAOQAOHAQOQPTMAPRATPTSAPPVQISTVQAPGTPIIAR-QVTP----- 518
QY 116 AQIGTTVTVTVPKSSVQSVAVPTSVVTPGKPLNTVT---TLKPSSILGASS-----TPS 168
DB 519 -----TTIIKQV---SQAQTTVPQSATLQSPGVQPOLVGLGAAQTASLGTATAVQTGTPQ 571
QY 169 NE-PNLKAENSAAVQINLSPMTLENVKKCNFLAMLIKACSGSQSPMGONVKKLYEQL 227
DB 572 RTVPGATTSSAAE-----TMENVKKCNFLSTLIKASSGKQSTETAAANVKELVQNL 625
QY 228 LDAKIEAEFTSRKLYVELKSSPQPHLPFLKKSVALRQLLPNSQSFIOQCQVQ-----TS 283

```
Db 626 LDKIEAEFTTSLYRELNSSSQPVLVFLKRLPALRLQTLTPDSAAFIQSQQQPPPTPS 685
QY 284 SDMWIATCTTWTSPVTTTSSSSQSEKSIIVSGATAPRTVSU 327
Db 686 Q-----ATTALTAVLSSSVQRTAGTAAVTYSALQPPVL 722

RESULT 10
PCT-US02-30474-1660
; Sequence 1660, Application PC/TUS0230474
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Ma, Yungqing
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Halsey-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 810CIP PCT
; CURRENT APPLICATION NUMBER: PCT/US02/30474
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/324,631
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3476
; SOFTWARE: pt_FL_genes Version 6.0
; SEQ ID NO 1660
; LENGTH: 1051
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-30474-1660

Query Match 27.7%; Score 491; DB 1; Length 1051;
Best Local Similarity 38.4%; Pred. No. 2.4e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTITIOFPANQLPPGTVLKSNGPLM 61
Db 382 PRATTSIGIRATLTPVLAPRLPOP-----PQNPTNIQ---NFQLPPGVLVRSENGQLL 432
QY 62 LVSPQQTWTR-----AETTSNITSRPVAPNPQVTKICTVPSNSSLIKKVVAVTPVKKL 115
Db 433 MI-PQALAQMAQAHAQAPQTTMAPRPATPTSPBPVQISTVQAPGTPIAR-QVTP---- 486
```

```
QY 116 AQIGTTVTTVPKPSVQSVAVPTSVVTVTPGKPLNTVT--TLKPSSLGASS-----TPS 168
Db 487 ----TTIIKQV---SOAQTTVPQSPATLQSRPGVQPOLVLGGAQAATSLGTATAVQTGPQ 539
QY 169 NE-PNLKAENSAVAQINLSPTMLENVKCKNFAMLIKACSGSQSPKMGQNYKLVQEL 227
Db 540 RTVPGATTSSAATE-----TMENVKCKNFELSTLIKASSGKQSTETAANYKLVQNL 593
QY 228 LDAKIEAEFTRLKLYVELKSSQPHLVLPVLKKSVALRQLLPNSQSFIOQCVOQ-----TS 283
Db 594 LDKIEAEFTTSLYRELNSSSQPVLVFLKRLPALRLQTLTPDSAAFIQSQQQPPPTPS 653
QY 284 SDMWIATCTTWTSPVTTTSSSSQSEKSIIVSGATAPRTVSU 327
Db 654 Q-----ATTALTAVLSSSVQRTAGTAAVTYSALQPPVL 690

RESULT 11
US-60-324-631-1665
; Sequence 1665, Application US/60324631
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Ma, Yungqing
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Halsey-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 810
; CURRENT APPLICATION NUMBER: US/60/324,631
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/04941
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31
```

;; PRIOR APPLICATION NUMBER: US 09/649,167
;; PRIOR FILING DATE: 2000-08-23
;; PRIOR APPLICATION NUMBER: PCT/US01/08631
;; PRIOR FILING DATE: 2001-03-30
;; PRIOR APPLICATION NUMBER: US 09/552,929
;; PRIOR FILING DATE: 2000-04-18
;; PRIOR APPLICATION NUMBER: US 09/770,160
;; PRIOR FILING DATE: 2001-01-26
;; PRIOR APPLICATION NUMBER: PCT/US01/08656
;; PRIOR FILING DATE: 2001-04-18
;; PRIOR APPLICATION NUMBER: US 09/577,408
;; PRIOR FILING DATE: 2000-05-18
;; PRIOR APPLICATION NUMBER: PCT/US01/14827
;; PRIOR FILING DATE: 2001-05-16
;; NUMBER OF SEQ ID NOS: 3334
;; SOFTWARE: pc_flg_genes Version 6.0
;; SEQ ID NO 1665
;; LENGTH: 1051
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-60-324-631-1665

Query Match 27.7%; Score 491; DB 27; Length 1051;
Best Local Similarity 38.4%; Pred. No. 2.4e-31;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

Qy 13 PKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPCTVLKNSGGLM 61
Db 382 PTATTSIGIRATLTPTVLAPRLPQ-----PQNPTNIO---NFQLPPGMVLVRSNGQLL 432
Qy 62 LVSPQQTVTR-----AETTSNISRPAVPANPOTVKICTVPSNSSOLIKKAVTPVKKL 115
Db 433 MI-PQALAQQAQAHAQPTMAPRPATPSAPPVQISTVQAGTPIIAR-QVTP----- 486
Qy 116 AQIGTVTVTVTPKPSVQSVAVPTSVTVTPGKPLNTVT--TLKPSSILGASS-----TPS 168
Db 487 ---TTIKQV---SQAGTVQPSATLQSPGVQPOLVGLGAAQTASLGATATVAGTGPQ 539
Qy 169 NE-PNLKAENSAAVQINLSPTMLENVKCKNFLAMLIKACSGSQSPENGMONVKKLVQL 227
Db 540 RTVPGATTTSAAETE-----TMENVKCKNFLSTLIKASSGKQSTETAANVKLVQNL 593
Qy 228 LDAKIEAEFFTRKLYELKSSQPQLHVPFLKKSVALRQLLPNSQSFIQCCVQO-----TS 283
Db 594 LDGKIEADFTSRLYRELNSSPQYLVFPFLKRSPLALRQLTPDAAFIQSQQQPPPTTS 653
Qy 284 SDMVIATCTTTTSPVTVTVSSSQSEKSIIVSGATAPRTVSV 327
Db 654 Q-----ATTALTAVLSSSVQRTAGTKTAATVTSALQPPVL 690

RESULT 12
PCT-US02-30474-3271
; Sequence 3271, Application PC/TUS0230474
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aigong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Ma, Yuning
; APPLICANT: Asundi, Vinod
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Haley-Vicente, Dana
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides

;; FILE REFERENCE: 810CIP PCT
;; CURRENT APPLICATION NUMBER: PCT/US02/30474
;; CURRENT FILING DATE: 2002-09-24
;; PRIOR APPLICATION NUMBER: US 60/324,631
;; PRIOR FILING DATE: 2001-09-24
;; PRIOR APPLICATION NUMBER: US 09/488,725
;; PRIOR FILING DATE: 2000-01-21
;; PRIOR APPLICATION NUMBER: US 09/552,317
;; PRIOR FILING DATE: 2000-04-25
;; PRIOR APPLICATION NUMBER: PCT/US00/35017
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: US 09/491,404
;; PRIOR FILING DATE: 2000-01-25
;; PRIOR APPLICATION NUMBER: PCT/US01/02623
;; PRIOR FILING DATE: 2001-01-25
;; PRIOR APPLICATION NUMBER: US 09/496,914
;; PRIOR FILING DATE: 2000-02-03
;; PRIOR APPLICATION NUMBER: US 09/560,875
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: PCT/US01/03800
;; PRIOR FILING DATE: 2001-02-05
;; PRIOR APPLICATION NUMBER: US 09/515,126
;; PRIOR FILING DATE: 2000-02-28
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 3476
;; SOFTWARE: pc_flg_genes Version 6.0
;; SEQ ID NO 3271
;; LENGTH: 1083
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(1083)
;; OTHER INFORMATION: Xaa = any amino acid or symbol as shown in the table 8 as set
PCT-US02-30474-3271

Query Match 27.0%; Score 479; DB 1; Length 1083;
Best Local Similarity 36.3%; Pred. No. 2.6e-30;
Matches 12; Conservative 50; Mismatches 97; Indels 76; Gaps 12;

Qy 13 PKVSSG-----PRLPAPQIVAVKAPNTTIOFPANLQLPCTVLKNSGGLM 61
Db 414 PTATTSIGIRATLTPTVLAPRLPQ-----PQNPTNIO---NFQLPPGMVLVRSNGQLL 464
Qy 62 LVSPQQTVTR-----AETTSNISRPAVPANPOTVKICTVPSNSSOLIKKAVTPVKKL 115
Db 465 MI-PQALAQQAQAHAQPTMAPRPATPSAPPVQISTVQAGTPIIAR-QVTP----- 522
Qy 116 AQIGTVTVTVTP-----KPSVQSVAVPTSVTVTP---GKPLNTVTVTLKPSL 161
Db 523 SRVQAQTTVQPSATLKRSPGSPQPOLVGLGAAQTASLGATATVAGTGPQRTVTP----- 575
Qy 162 GASSTPSNEPNLKAENSAAVQINLSPTMLENVKCKNFLAMLIKACSGSQSPENGMONV 221
Db 576 GATTTS-----VARETMENVKCKNFLKASSGKQSTETAANVK 619
Qy 222 KLVQLDLAKIEAEFFTRKLYELKSSQPQLHVPFLKKSVALRQLLPNSQSFIQCCVQO 281
Db 620 ELVQNLGKIEADFTSRLYRELNSSPQYLVFPFLKRSPLALRQLTPDAAFIQSQQQ 679
Qy 282 ---TSSDMVIATCTTTTSPVTVTVSSSQSEKSIIVSGATAPRTVSV 327
Db 680 PPTTSQ-----ATTALTAVLSSSVQRTAGTKTAATVTSALQPPVL 722

RESULT 13
US-60-167-217-11401
; Sequence 11401, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL000152
CURRENT APPLICATION NUMBER: US/60/167,217
CURRENT FILING DATE: 1999-11-24
NUMBER OF SEQ ID NOS: 23195
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11401
LENGTH: 899
TYPE: PRT
ORGANISM: Drosophila
US-60-167-217-11401

Query Match 12.3%; Score 218; DB 27; Length 899;
Best Local Similarity 24.9%; Pred. No. 1.6e-08;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIVAVKAPNTTIOFPANLQPPGTVLKSNLGLM-----LVKTDNGFQLLRVGTGTTGTTTQTITNSNS 162
DB 111 POSPSITLSTLTGQTPA-----LLKDVAVTPVKKLAQIGTTVVTTVP 127
QY 77 NITSPAVPANPQTVKICVTPNSSQ-----LIKKVAVTPVKKLAQIGTTVVTTVP 127
DB 163 NITSTTNHPTTQ-IRLQTPAAASMTNTATSNIIIVSVASSGYANSQPPHLLTQNAQ 221
QY 128 KPSSVQSVAVPTSVVTPGKPLNTVTLKPSLSGASSTPSNEPNLKAENSAVQINLSP 187
DB 222 APQLPQITQIOTIPAOQSQOQVNNVSSAGGTATAVSSTA-----ATT 265
QY 188 TMLENVK-KCKNFLAMLIKACSGSPGEMGVKLVQQLDADAKIEAEFTKLYVELK 246
DB 266 TOGNTKECKRFLANLIEL--STREPKEPVKNVRLTQELVNAVPEEFCDRLERLN 323
QY 247 SSPQHLVPFLKKSVALRQL-----LPNSQSFQ--- 276
DB 324 ASPQCLIGFLKSLPLLRQALYKELVIEGKPPQHVGLGLAGLSQQLPKIQAIIRPIG 383
QY 277 ----OCVQOTSSDMVIACTTTTSPVTTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 384 PSQTTTIGTQVRMI--TPNALGTTPRTTIGHTTISKOPPN---IRLPTAPRLVNTGGIRT 438
QY 330 LNPLAGPVGAKAGVTVLHSGVPTAA 354
DB 439 QIP-SLQVPGQANIVQIR--GPQHA 460

RESULT 14

US-60-173-464-9223
Sequence 9223, Application US/60173464

GENERAL INFORMATION:
APPLICANT: LI, Peter W.D.
TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
FILE REFERENCE: CL000173
CURRENT APPLICATION NUMBER: US/60/173,464
CURRENT FILING DATE: 1999-12-29
NUMBER OF SEQ ID NOS: 30269
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9223
LENGTH: 921
TYPE: PRT
ORGANISM: Drosophila
US-60-173-464-9223

Query Match 12.3%; Score 218; DB 27; Length 921;
Best Local Similarity 24.9%; Pred. No. 1.6e-08;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIVAVKAPNTTIOFPANLQPPGTVLKSNLGLM-----LVKTDNGFQLLRVGTGTTGTTTQTITNSNS 184
DB 133 POSPSITLSTLTGQTPA-----LIKKVAVTPVKKLAQIGTTVVTTVP 127
QY 77 NITSPAVPANPQTVKICVTPNSSQ-----LIKKVAVTPVKKLAQIGTTVVTTVP 127

DB 185 NITSTTNHPTTQ-IRLQTPAAASMTNTATSNIIIVSVASSGYANSQPPHLLTQNAQ 243
QY 128 KPSSVQSVAVPTSVVTPGKPLNTVTLKPSLSGASSTPSNEPNLKAENSAVQINLSP 187
DB 244 APQLPQITQIOTIPAOQSQOQVNNVSSAGGTATAVSSTA-----ATT 287
QY 188 TMLENVK-KCKNFLAMLIKACSGSPGEMGVKLVQQLDADAKIEAEFTKLYVELK 246
DB 288 TOGNTKECKRFLANLIEL--STREPKEPVKNVRLTQELVNAVPEEFCDRLERLN 345
QY 247 SSPQHLVPFLKKSVALRQL-----LPNSQSFQ--- 276
DB 346 ASPQCLIGFLKSLPLLRQALYKELVIEGKPPQHVGLGLAGLSQQLPKIQAIIRPIG 405
QY 277 ----OCVQOTSSDMVIACTTTTSPVTTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGTQVRMI--TPNALGTTPRTTIGHTTISKOPPN---IRLPTAPRLVNTGGIRT 460
QY 330 LNPLAGPVGAKAGVTVLHSGVPTAA 354
DB 461 QIP-SLQVPGQANIVQIR--GPQHA 482

RESULT 15

US-60-173-464-20611
Sequence 20611, Application US/60173464

GENERAL INFORMATION:
APPLICANT: LI, Peter W.D.
TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
FILE REFERENCE: CL000173
CURRENT APPLICATION NUMBER: US/60/173,464
CURRENT FILING DATE: 1999-12-29
NUMBER OF SEQ ID NOS: 30269
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20611
LENGTH: 921
TYPE: PRT
ORGANISM: Drosophila
US-60-173-464-20611

Query Match 12.3%; Score 218; DB 27; Length 921;
Best Local Similarity 24.9%; Pred. No. 1.6e-08;
Matches 96; Conservative 58; Mismatches 141; Indels 90; Gaps 14;

QY 25 POIVAVKAPNTTIOFPANLQPPGTVLKSNLGLM-----LVKTDNGFQLLRVGTGTTGTTTQTITNSNS 184
DB 133 POSPSITLSTLTGQTPA-----LIKKVAVTPVKKLAQIGTTVVTTVP 127
QY 77 NITSPAVPANPQTVKICVTPNSSQ-----LIKKVAVTPVKKLAQIGTTVVTTVP 127
DB 185 NITSTTNHPTTQ-IRLQTPAAASMTNTATSNIIIVSVASSGYANSQPPHLLTQNAQ 243
QY 128 KPSSVQSVAVPTSVVTPGKPLNTVTLKPSLSGASSTPSNEPNLKAENSAVQINLSP 187
DB 244 APQLPQITQIOTIPAOQSQOQVNNVSSAGGTATAVSSTA-----ATT 287
QY 188 TMLENVK-KCKNFLAMLIKACSGSPGEMGVKLVQQLDADAKIEAEFTKLYVELK 246
DB 288 TOGNTKECKRFLANLIEL--STREPKEPVKNVRLTQELVNAVPEEFCDRLERLN 345
QY 247 SSPQHLVPFLKKSVALRQL-----LPNSQSFQ--- 276
DB 346 ASPQCLIGFLKSLPLLRQALYKELVIEGKPPQHVGLGLAGLSQQLPKIQAIIRPIG 405
QY 277 ----OCVQOTSSDMVIACTTTTSPVTTVSSSQSEKSIIVSGATAPRTVS---VQT 329
DB 406 PSQTTTIGTQVRMI--TPNALGTTPRTTIGHTTISKOPPN---IRLPTAPRLVNTGGIRT 460
QY 330 LNPLAGPVGAKAGVTVLHSGVPTAA 354
DB 461 QIP-SLQVPGQANIVQIR--GPQHA 482

Search completed: February 16, 2003, 22:18:31
Job time : 133.688 secs

This Page Blank (uspto)

; Sequence 63929, Application US/09724676A

; Sequence 63929, Application US/09724676A

```

; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 63929
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63929

Query Match      27.7%; Score 491; DB 5; Length 679;
Best Local Similarity 38.4%; Pred. No. 4,8e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14

Y 13 PEKVSSG-----RPLPAPQIVAAKAPNTTIOFPANQLPPGCVILKSNCGPLM 61
      |::||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 286 PLATISGIRATLTPVILAPRLQP-----PQNPINIO---NFQLPQGVILVRSNGQLL 336
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Y 62 LVSPQQTIVTR-----AETTSNITSRPAPVAPNPQTVKICTVPSNSQLIKKVAVTPVKTL 115
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 337 MI-PQALALQMOQAHAQOQTMMAPRAPPTSPAPVQISTVQAQCPRIIAR-QVTP----- 390
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Y 116 AQIGTTVTVTKRPSSVQSVANPTSVTVYTPGKPLNTVY--TLKPSLCASS-----TPS 168
      391 ----TTLIKQV---SQQQTTPQPSATLQSPGVQPOLVIGAAQVATSLCTATAVQTGTPQ 443
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Y 169 NE-PMLKAENSAAYQINSPMLLENVKKCKKNFLAMLIKLAGSGSGSPKGGONVKKLYEQL 227
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 444 RVPKPAITTSNAATE-----TMEVKKCKKNLSTILIKLASSGKOSTETAAANKELVONL 497
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Y 228 IDAKTEAEETFRKLYELKSSPQPHLVPFLKRSVALROLLEPNSSQFIOQVQO---TS 283
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 498 LDGKTEAEDFTSRILKRELNSPQYLVPLPKRLSLPALROLTDPDSAAFIOQSQQQPPPTTS 557
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Y 284 SMMVATATCTTYTTSPPVYTTTVSSSGSEKSIYSGATPRATSV 327
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 558 Q-----ATTALTAVLVSSSVQRTAGKTAATVTSALQPPVLT 594
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
US-09-724-676-63927
; Sequence 63927, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 63927
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63927

Query Match      27.7%; Score 491; DB 5; Length 709;
Best Local Similarity 38.4%; Pred. No. 5e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14

Y 13 PEKVSSG-----RPLPAPQIVAAKAPNTTIOFPANQLPPGCVILKSNCGPLM 61
      |::||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 286 PLATISGIRATLTPVILAPRLQP-----PQNPINIO---NFQLPQGVILVRSNGQLL 336
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Y 62 LVSPQQTIVTR-----AETTSNITSRPAPVAPNPQTVKICTVPSNSQLIKKVAVTPVKTL 115
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 337 MI-PQALALQMOQAHAQOQTMMAPRAPPTSPAPVQISTVQAQCPRIIAR-QVTP----- 390
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Y 116 AQIGTTVTVTKRPSSVQSVANPTSVTVYTPGKPLNTVY--TLKPSLCASS-----TPS 168

```

```

Db      391  ----TTIIKQV---SQAQITVQPSAATLQNSPCGVQQLVLGGAQRTASLSGTAAVQNGCTGQ  443
OY      169  NE-PNLKKEANSAAVOINLSPTMLENVKKCKNPLAMLIKTIACSGSGSPENKGAVYKKEVLQ  227
Db      444  RTVPATTTSSAATE-----TMENVKKCKNPLSTLIKTIACSGSGSTETAAVKKELVQNL  497
OY      228  LDKATKEAEFFTKRLVYELKSSPPQPHLYPPLKKSVAALRDLNNSGQSFIOCCVQO-----TS  283
Db      498  LDGKLEAEDEFTSLRELNLSPPQPLVPLPKRSPLALRQLTDPDSAAFIOQSOQOPPPPTS  557
OY      284  SDMWATCTTCTVTTSPVYTTVYSSSGSEKSIIVSGATAPRTVS  327
Db      558  Q-----ATTALTAIVLSSSVQRTAGKTAATVTSALQPPVUL  594

RESULT 4
US-09-724-676A-63927
Sequence 63927, Application US/09724676A
GENERAL INFORMATION:
APPLICANT: Comugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Comugen
CURRENT APPLICATION NUMBER: US/09/724,676A
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63927
LENGTH: 709
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676A-63927

Query Match      27.7%; Score 491; DB 5; Length 709;
Best Local Similarity 38.4%; Pident.No. 5e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14.

OY      13  PKVSSG-----PRLEAPQIVAAKAPNTTTTIOPPANLQLPFGVILIKNSGLM  61
Db      286  PTAATSGIRATLTPVLAERLPQ-----PONPTNIQ--NFOLEPGMVLVRESENGQL  336
OY      62  LVSPQQTWR-----AETTSNITSRPAPAPAPQVYKICTVNSSQLKKVAVTPYKRL  115
Db      337  MI-PQALAAQMOQAHAQOQTMMAPRPATVTSAPPQISTVOAPGPIIAR-QVTP---  390
OY      116  AQIGTIVTVTPKPSVSVAVPTSVYTTPGKPLNTVY--TLKPSLSGASS-----TPS  168
Db      391  ----TTIIKQV---SQAQITVQPSAATLQNSPCGVQQLVLGGAQRTASLSGTAAVQNGCTGQ  443
OY      169  NE-PNLKKEANSAAVOINLSPTMLENVKKCKNPLAMLIKTIACSGSGSPENKGAVYKKEVLQ  227
Db      444  RTVPATTTSSAATE-----TMENVKKCKNPLSTLIKTIACSGSGSTETAAVKKELVQNL  497
OY      228  LDKATKEAEFFTKRLVYELKSSPPQPHLYPPLKKSVAALRDLNNSGQSFIOCCVQO-----TS  283
Db      498  LDGKLEAEDEFTSLRELNLSPPQPLVPLPKRSPLALRQLTDPDSAAFIOQSOQOPPPPTS  557
OY      284  SDMWATCTTCTVTTSPVYTTVYSSSGSEKSIIVSGATAPRTVS  327
Db      558  Q-----ATTALTAIVLSSSVQRTAGKTAATVTSALQPPVUL  594

RESULT 5
US-09-724-676-63920
Sequence 63920, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Comugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Comugen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2

```



```

Db 433 MI-PQALAQOAHQAPOTTMARPATPTSAPPVOISTVOAECTPIIAR-QVTP----- 486
Oy 116 AOGITVTVTPKPSVSVAVPPTSIVTTPGKPLNTVT--TLKPSIGASS-----TPS 168
Db 487 -----TTIIKOV-----SQAOTVQPSATLQSPGVOPQVLGGAAGTASIGTATAVOTGTPO 539
Oy 169 NE-PLKAKENSAVOINLSPTMLENVKCKNFMLILKLACSGSOPMGONVKKLVEQL 227
Db 540 RTVPGATTSSAATE-----TMENVKCKNFSLTLKLACSGSKOSTETANVKELVONL 593
Oy 228 LDKATEAEFTRLKVELKSSPOPHLPFLKKSVALROLPLNSOSFTIOQCVQO-----TS 283
Db 594 LDGKTEADEFTSRLYRELNSSPOPLVPEFLKRSLEPALROLPLPDSAAFTIOQSOQOOPPPPTS 653
Oy 284 SDMWIATCTTIVTSPVTVTVSSOSSEKSIIVSGATAPRTVS 327
Db 654 Q-----ATTALTAIVLVSSSVORTAGKTAATVTSALOPPVLSL 690

RESULT 12
US-09-724-676A-63921
; Sequence 63921, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63921
; LENGTH: 805
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63921

Query Match
Best Local Similarity 38.4%; Score 491; DB 5; Length 805;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

Oy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTFOPANLQLPCTVLIKSNSGPLM 61
Db 382 PTATTSGRATLTPVLAPRLPQP-----PONPNINQ-----NFOLPQGMVLVRSENGQL 432
Oy 62 LVSPQOQVTR-----AETTSNITSRPVAVPANPQTVKICITVPNSSQOLIKKVAATVPVKL 115
Db 433 MI-PQALAQOAHQAPOTTMARPATPTSAPPVOISTVOAECTPIIAR-QVTP----- 486
Oy 116 AOGITVTVTPKPSVSVAVPPTSIVTTPGKPLNTVT--TLKPSIGASS-----TPS 168
Db 487 -----TTIIKOV-----SQAOTVQPSATLQSPGVOPQVLGGAAGTASIGTATAVOTGTPO 539
Oy 169 NE-PLKAKENSAVOINLSPTMLENVKCKNFMLILKLACSGSOPMGONVKKLVEQL 227
Db 540 RTVPGATTSSAATE-----TMENVKCKNFSLTLKLACSGSKOSTETANVKELVONL 593
Oy 228 LDKATEAEFTRLKVELKSSPOPHLPFLKKSVALROLPLNSOSFTIOQCVQO-----TS 283
Db 594 LDGKTEADEFTSRLYRELNSSPOPLVPEFLKRSLEPALROLPLPDSAAFTIOQSOQOOPPPPTS 653
Oy 284 SDMWIATCTTIVTSPVTVTVSSOSSEKSIIVSGATAPRTVS 327
Db 654 Q-----ATTALTAIVLVSSSVORTAGKTAATVTSALOPPVLSL 690

RESULT 13
US-09-724-676-63926
; Sequence 63926, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63926
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63926

Query Match
Best Local Similarity 38.4%; Score 491; DB 5; Length 807;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

```

```

; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63926
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-63926

Query Match
Best Local Similarity 38.4%; Score 491; DB 5; Length 807;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

Oy 13 PPKVSSG-----PRLPAPQIVAVKAPNTTTFOPANLQLPCTVLIKSNSGPLM 61
Db 414 PTATTSGRATLTPVLAPRLPQP-----PONPNINQ-----NFOLPQGMVLVRSENGQL 464
Oy 62 LVSPQOQVTR-----AETTSNITSRPVAVPANPQTVKICITVPNSSQOLIKKVAATVPVKL 115
Db 465 MI-PQALAQOAHQAPOTTMARPATPTSAPPVOISTVOAECTPIIAR-QVTP----- 518
Oy 116 AOGITVTVTPKPSVSVAVPPTSIVTTPGKPLNTVT--TLKPSIGASS-----TPS 168
Db 519 -----TTIIKOV-----SQAOTVQPSATLQSPGVOPQVLGGAAGTASIGTATAVOTGTPO 571
Oy 169 NE-PLKAKENSAVOINLSPTMLENVKCKNFMLILKLACSGSOPMGONVKKLVEQL 227
Db 572 RTVPGATTSSAATE-----TMENVKCKNFSLTLKLACSGSKOSTETANVKELVONL 625
Oy 228 LDKATEAEFTRLKVELKSSPOPHLPFLKKSVALROLPLNSOSFTIOQCVQO-----TS 283
Db 626 LDGKTEADEFTSRLYRELNSSPOPLVPEFLKRSLEPALROLPLPDSAAFTIOQSOQOOPPPPTS 685
Oy 284 SDMWIATCTTIVTSPVTVTVSSOSSEKSIIVSGATAPRTVS 327
Db 686 Q-----ATTALTAIVLVSSSVORTAGKTAATVTSALOPPVLSL 722

RESULT 14
US-09-724-676A-63926
; Sequence 63926, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63926
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-63926

Query Match
Best Local Similarity 38.4%; Score 491; DB 5; Length 807;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

```

Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

```
OY 13 PRVSSG-----PRLPAQIVAAKAPTTTTIOFPANQLPFGVLIKNSGPI 61
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 318 PRATSGIRATLPTVLARLP-----PQNPNIQ---NQLPFGVLIKNSGQ 368
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 62 LVSPOQVTR-----AETTSNITSRPVAVPANDQVTKICTVPNSSQOLIKKVA 115
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 369 MI-PQALAQMOQAHAQOQTMAPRATPTSAPVOISTVQAPGPIIAR-OVTP 422
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 116 AQGTGTVVTVTPKSSVQAVPTSVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 423 ----TTIIKOV---SOAQTVQPSATLORS PGVQPOLVIGGAQOTASLGATAVQGTPO 475
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 169 NE-PNLKENSAAVQINLSPTMLENKCKNFMILIKLACSSGSGSPKMGQNYKLE 227
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 476 RTVPGATTTSSATE-----TMENVKCKNFMILIKLACSSGSGSPKMGQNYKLE 529
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 228 LDKIAEERTKRLVLELKSSPOPHLVPLKRSVALROLLPNSQSFIOOCVOO---TS 283
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 530 LDKIAEEDFTSRLYRELNSSPOPIVLPFLKRSPLROLTPDSAPAFIOOSQO 589
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 284 SDMVATCTTTVTSPVTTVSSSGSEKSIIVSGATAPRTSV 327
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 590 Q-----ATTALTAVLSSSVQRTAGTAATVTSALQPPVLSL 626
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 9

US-09-724-676-63923
Sequence 63923, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724, 676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63923
LENGTH: 775
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-63923

Query Match 27.7%; Score 491; DB 5; Length 775;
Best Local Similarity 38.4%; Pred. No. 5, 5e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

```
OY 13 PRVSSG-----PRLPAQIVAAKAPTTTTIOFPANQLPFGVLIKNSGPI 61
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 382 PRATSGIRATLPTVLARLP-----PQNPNIQ---NQLPFGVLIKNSGQ 432
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 62 LVSPOQVTR-----AETTSNITSRPVAVPANDQVTKICTVPNSSQOLIKKVA 115
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 433 MI-PQALAQMOQAHAQOQTMAPRATPTSAPVOISTVQAPGPIIAR-OVTP 486
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 116 AQGTGTVVTVTPKSSVQAVPTSVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 487 ----TTIIKOV---SOAQTVQPSATLORS PGVQPOLVIGGAQOTASLGATAVQGTPO 539
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 169 NE-PNLKENSAAVQINLSPTMLENKCKNFMILIKLACSSGSGSPKMGQNYKLE 227
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 540 RTVPGATTTSSATE-----TMENVKCKNFMILIKLACSSGSGSPKMGQNYKLE 593
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 228 LDKIAEERTKRLVLELKSSPOPHLVPLKRSVALROLLPNSQSFIOOCVOO---TS 283
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 594 LDKIAEEDFTSRLYRELNSSPOPIVLPFLKRSPLROLTPDSAPAFIOOSQO 653
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 284 SDMVATCTTTVTSPVTTVSSSGSEKSIIVSGATAPRTSV 327
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 654 Q-----ATTALTAVLSSSVQRTAGTAATVTSALQPPVLSL 690
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 10

US-09-724-676A-63923
Sequence 63923, Application US/09724676A
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724, 676A
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63923
LENGTH: 775
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676A-63923

Query Match 27.7%; Score 491; DB 5; Length 775;
Best Local Similarity 38.4%; Pred. No. 5, 5e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

```
OY 13 PRVSSG-----PRLPAQIVAAKAPTTTTIOFPANQLPFGVLIKNSGPI 61
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 382 PRATSGIRATLPTVLARLP-----PQNPNIQ---NQLPFGVLIKNSGQ 432
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 62 LVSPOQVTR-----AETTSNITSRPVAVPANDQVTKICTVPNSSQOLIKKVA 115
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 433 MI-PQALAQMOQAHAQOQTMAPRATPTSAPVOISTVQAPGPIIAR-OVTP 486
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 116 AQGTGTVVTVTPKSSVQAVPTSVTVTPGKPLNTVT--TLKPSLGLASS-----TPS 168
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 487 ----TTIIKOV---SOAQTVQPSATLORS PGVQPOLVIGGAQOTASLGATAVQGTPO 539
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 169 NE-PNLKENSAAVQINLSPTMLENKCKNFMILIKLACSSGSGSPKMGQNYKLE 227
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 540 RTVPGATTTSSATE-----TMENVKCKNFMILIKLACSSGSGSPKMGQNYKLE 593
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 228 LDKIAEERTKRLVLELKSSPOPHLVPLKRSVALROLLPNSQSFIOOCVOO---TS 283
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 594 LDKIAEEDFTSRLYRELNSSPOPIVLPFLKRSPLROLTPDSAPAFIOOSQO 653
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 284 SDMVATCTTTVTSPVTTVSSSGSEKSIIVSGATAPRTSV 327
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 654 Q-----ATTALTAVLSSSVQRTAGTAATVTSALQPPVLSL 690
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 11

US-09-724-676-63921
Sequence 63921, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724, 676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 63921
LENGTH: 805
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-63921

Query Match 27.7%; Score 491; DB 5; Length 805;
Best Local Similarity 38.4%; Pred. No. 5, 7e-21;
Matches 132; Conservative 52; Mismatches 96; Indels 64; Gaps 14;

```
OY 13 PRVSSG-----PRLPAQIVAAKAPTTTTIOFPANQLPFGVLIKNSGPI 61
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 382 PRATSGIRATLPTVLARLP-----PQNPNIQ---NQLPFGVLIKNSGQ 432
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 62 LVSPOQVTR-----AETTSNITSRPVAVPANDQVTKICTVPNSSQOLIKKVA 115
      | : | | | | | | | | | | | | | | | | | | | | | | | | | |
```


XX Dikstein R, Tjian R;
 XX WPI; 1998-109818/10.
 DR P-PSDB; AAW31494.
 XX
 XX DNA encoding human tata-binding protein associated factor - for
 XX producing recombinant protein
 XX
 XX Claim 7; Col 11-18; 12pp; English.
 XX
 XX This cDNA sequence encodes a human tata-binding protein associated
 CC factor, hTAFII105, isolated from Daudi cell nuclear extracts. Tightly
 CC associated subunits (TAF's) are components of the transcription factor
 CC TFIID and are thought to mediate transcriptional activation. This encoded
 CC protein may be produced recombinantly from transformed host cells or
 CC purified from human cells. hTAFII105 specific binding agents such as
 CC specific antibodies could be used for diagnosis (e.g. genetic
 CC hybridisation screens for hTAFII105 transcripts), therapy (e.g. gene
 CC therapy to modulate hTAFII105 gene expression) and in the
 CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
 CC cell specific activators or other transcriptional regulators).
 XX
 SQ Sequence 2556 BP; 797 A; 614 C; 514 G; 631 T; 0 other;
 Query Match 99.1%; Score 2534; DB 19; Length 2556;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
 QY 1 GGGACCTGGTGACCAAAAGTGGCTCGGTCAGCGCCCTCCCTAAAGTCAGAGCGGCCCT 60
 DB 1 GGGACCTGGTGACCAAAAGTGGCTCGGTCAGCGCCCTCCCTAAAGTCAGAGCGGCCCT 60
 QY 61 AGGTCGCTGCTCCTCAGATAGTCGCCGTGAAGCCCCCAACACACAGCAATFCCAGTTT 120
 DB 61 AGGTCGCTGCTCCTCAGATAGTCGCCGTGAAGCCCCCAACACACAGCAATFCCAGTTT 120
 QY 121 CTGCTAAATTTGAGCTTCTCCAGGAACCGTTTTCATTAAAGTAACAGTGTCCGTTG 180
 DB 121 CTGCTAAATTTGAGCTTCTCCAGGAACCGTTTTCATTAAAGTAACAGTGTCCGTTG 180
 QY 181 ATGTTGGTATCTCCTCAGCAAACTGTAAAGAGCGGAGACACCAAGTAACATACCTCA 240
 DB 181 ATGTTGGTATCTCCTCAGCAAACTGTAAAGAGCGGAGACACCAAGTAACATACCTCA 240
 QY 241 AGGCCAGCAGTACCAGCAATCTCTCAACAGTCAAAATCTGACGTGCCGAACTTACG 300
 DB 241 AGGCCAGCAGTACCAGCAATCTCTCAACAGTCAAAATCTGACGTGCCGAACTTACG 300
 QY 301 TCACAAATTAATCAAGAAAGTGGCAGTGACACCTGTGTTAAAAATTTGGCACAATAGGA 360
 DB 301 TCACAAATTAATCAAGAAAGTGGCAGTGACACCTGTGTTAAAAATTTGGCACAATAGGA 360
 QY 361 ACTGGGTAAACCACTGTTCCGAAGCCTTCCCTCAGTACAATCTGTGGCTGTGCCAACCA 420
 DB 361 ACTGGGTAAACCACTGTTCCGAAGCCTTCCCTCAGTACAATCTGTGGCTGTGCCAACCA 420
 QY 421 RCTGCTCAGAGTACTCTCTGGAAGCCATTGAATAGTACTACCTACCTGAGCCCTTCA 480
 DB 421 -GTGCTCAGAGTACTCTCTGGAAGCCATTGAATAGTACTACCTGAGCCCTTCA 479
 QY 481 TTTGGAGCATCATCACCTCTTCAATGAGCCCAATCTTAAAGCAGAGCAACTCAGCAGC 540
 DB 480 TTTGGAGCATCATCACCTCTTCAATGAGCCCAATCTTAAAGCAGAGCAACTCAGCAGC 539
 QY 541 TGTTCAGATTAATCTTCTCCGACAAATGCTAGAAAATGTGAAGAAATGCAAGAACTTCT 600
 DB 540 TGTTCAGATTAATCTTCTCCGACAAATGCTAGAAAATGTGAAGAAATGCAAGAACTTCT 599
 QY 601 TGCATGTTAAATAACTAGCATGTAGTGATACAGTCCCTGAAATGGGGCAAAATGT 660
 DB 600 TGCATGTTAAATAACTAGCATGTAGTGATACAGTCCCTGAAATGGGGCAAAATGT 659

QY 661 GAAGAAGCTGGTGGAAACAACCTTTTGGATGCAAAAATCGAAGCAGAGAAATTTACTAGGAA 720
 DB 660 GAAGAAGCTGGTGGAAACAACCTTTTGGATGCAAAAATCGAAGCAGAGAAATTTACTAGGAA 719
 QY 721 ACTGTATGTTGAACCTCAAGTCTTCACTCAGCTCCTCAGCTGGTTCTTTCTTAAGAAAG 780
 DB 720 ACTGTATGTTGAACCTCAAGTCTTCACTCAGCTCCTCAGCTGGTTCTTTCTTAAGAAAG 779
 QY 781 CGTGGTTGCTTTACGACAACTTCTGCTTAACCTCCAGAGCTTTCATCAGCAATGTGTTC 840
 DB 780 CGTGGTTGCTTTACGACAACTTCTGCTTAACCTCCAGAGCTTTCATCAGCAATGTGTTC 839
 QY 841 GCAGACTTCTAGTGACATGGTCAATGCTACTACTACAAAGTAACTTCTCTCTCTCT 900
 DB 840 GCAGACTTCTAGTGACATGGTCAATGCTACTACTACAAAGTAACTTCTCTCTCTCT 899
 QY 901 GGTGACAACTACAGTCTCTCAAGCCAGTCTGAAAAGTCAATTTATGTTCTCGAGCAAC 960
 DB 900 GGTGACAACTACAGTCTCTCAAGCCAGTCTGAAAAGTCAATTTATGTTCTCGAGCAAC 959
 QY 961 AGCACCAGAACTGTGTCTAGTGCATACTTTGAACCCACTTGTGTCAGTGGGAGCAAA 1020
 DB 960 AGCACCAGAACTGTGTCTAGTGCATACTTTGAACCCACTTGTGTCAGTGGGAGCAAA 1019
 QY 1021 AGCTGAGTGTGACACTTCAATTTCTGGGCCCACTGCTGCAACAGGAGGAACAACAGC 1080
 DB 1020 AGCTGAGTGTGACACTTCAATTTCTGGGCCCACTGCTGCAACAGGAGGAACAACAGC 1079
 QY 1081 TGAACCTGGTTTGTCTTTCAGACTTCAAAACCACTTGTGACATCTGTGCAACACAGTGAC 1140
 DB 1080 TGAACCTGGTTTGTCTTTCAGACTTCAAAACCACTTGTGACATCTGTGCAACACAGTGAC 1139
 QY 1141 CAGGCTCTCACTGCAACCTGAAAAGCCAGTGTCTCTGGAACACAGTAACTGTCCCT 1200
 DB 1140 CAGGCTCTCACTGCAACCTGAAAAGCCAGTGTCTCTGGAACACAGTAACTGTCCCT 1199
 QY 1201 TCCAGCAGTAACTTTTGGAGAACTTTCAGTGCAGCTATTTGCTTCCATCTGTGAAAC 1260
 DB 1200 TCCAGCAGTAACTTTTGGAGAACTTTCAGTGCAGCTATTTGCTTCCATCTGTGAAAC 1259
 QY 1261 TGTGTTTCTCTCTGCTGGGACCACTCTGCAAGCCCTGTTATTTGGGACTCCAGTTCAAAT 1320
 DB 1260 TGTGTTTCTCTCTGCTGGGACCACTCTGCAAGCCCTGTTATTTGGGACTCCAGTTCAAAT 1319
 QY 1321 CAAACTTGGCCAGCGCGGCCCTGCTCTTTCACAAACAGCTGGGATTTCCAAACAGGAGTTC 1380
 DB 1320 CAAACTTGGCCAGCGCGGCCCTGCTCTTTCACAAACAGCTGGGATTTCCAAACAGGAGTTC 1379
 QY 1381 AAGCAAGCAACTATTTCTCATTTCTTTCACCTAGTTCAGGAGCCCTTCAGGAGGCAATGAAA 1440
 DB 1380 AAGCAAGCAACTATTTCTCATTTCTTTCACCTAGTTCAGGAGCCCTTCAGGAGGCAATGAAA 1439
 QY 1441 ACAAGTGACCAAAATTTCACTTCCCTCAACATTTGACCAATTCAGAAATGTGGACAGAGAC 1500
 DB 1440 ACAAGTGACCAAAATTTCACTTCCCTCAACATTTGACCAATTCAGAAATGTGGACAGAGAC 1499
 QY 1501 GATGCCAGTGACACCAATTAATACCTACTAGTCACTTCTTCCAGCTTCCATTTCAAGCA 1560
 DB 1500 GATGCCAGTGACACCAATTAATACCTACTAGTCACTTCTTCCAGCTTCCATTTCAAGCA 1559
 QY 1561 AATTACCTCTGCTGGAATAAATTTCTCTCACTTCAAGCATCTCTCTACTCAGAAAATA 1620
 DB 1560 AATTACCTCTGCTGGAATAAATTTCTCTCACTTCAAGCATCTCTCTACTCAGAAAATA 1618
 QY 1621 GAATAAAGAGATGTAAACATCATGCTTCCGAGATGAGGATGACATCAATGATGTGACTT 1680
 DB 1619 GAATAAAGAGATGTAAACATCATGCTTCCGAGATGAGGATGACATCAATGATGTGACTT 1678
 QY 1681 CTATGCCAGGGTCAACCTTTAATGAAGAAAATGCGCTGCATCTTAGCAACAACTCTGAAT 1740
 DB 1679 CTATGCCAGGGTCAACCTTTAATGAAGAAAATGCGCTGCATCTTAGCAACAACTCTGAAT 1738
 QY 1741 TGGTTGGCAGTCTCATTCTAGTCAATGTAAAGATGAACCAATTTCTTTTATTTGGAGCTCTAC 1800

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:02:45 ; Search time 513 Seconds
(without alignments)
11229.263 Million cell updates/sec

Title: US-09-763-909-1

Perfect score: 2558

Sequence: 1 gggaccctggtagccaaagt.....ctttattacttacctat 2558

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 2534 | 99.1 | 2556 | 19 | Human hTAFII105 cd |
| 2 | 2534 | 99.1 | 2556 | 21 | Transcription fact |
| 3 | 2483.6 | 97.1 | 3468 | 23 | DNA encoding novel |
| 4 | 338.2 | 13.2 | 3603 | 17 | Human TATA-binding |
| 5 | 338.2 | 13.2 | 3603 | 18 | TATA-binding prote |
| 6 | 337.2 | 13.2 | 2214 | 15 | TATA-binding prote |
| 7 | 143.8 | 5.6 | 4520 | 23 | Drosophila melanog |
| 8 | 143.8 | 5.6 | 4520 | 23 | Drosophila melanog |
| 9 | 142.2 | 5.6 | 4615 | 15 | TATA-binding prote |

| | | | | | | |
|----|-------|-----|--------|----|----------|--------------------|
| 10 | 142.2 | 5.6 | 4615 | 17 | AA742210 | Drosophila TATA-bl |
| 11 | 142.2 | 5.6 | 4615 | 18 | AA779604 | TATA-binding prote |
| 12 | 135.8 | 5.3 | 2307 | 24 | ABK32842 | DNA encoding human |
| 13 | 60 | 2.3 | 60 | 24 | ABN33770 | Human spliced tran |
| 14 | 58.6 | 2.3 | 9525 | 23 | ABL09422 | Drosophila melanog |
| 15 | 58.6 | 2.3 | 10083 | 23 | ABL18476 | Drosophila melanog |
| 16 | 55.8 | 2.2 | 486 | 22 | AA775507 | Polyglutamine trac |
| 17 | 54 | 2.1 | 8201 | 21 | AAA88864 | Human dentin sialo |
| 18 | 54 | 2.1 | 8201 | 24 | ABQ73537 | Human dentin sialo |
| 19 | 46.2 | 1.8 | 2439 | 23 | ABL18865 | Drosophila melanog |
| 20 | 46.2 | 1.8 | 4439 | 23 | ABL18864 | Drosophila melanog |
| 21 | 42.4 | 1.7 | 5120 | 22 | AA846677 | DNA sequence of hu |
| 22 | 41.4 | 1.6 | 452 | 22 | AAK79268 | Human immune/haema |
| 23 | 41.4 | 1.6 | 583 | 22 | AAK79267 | Human immune/haema |
| 24 | 41 | 1.6 | 2169 | 21 | AAA30203 | Human RING finger |
| 25 | 40.8 | 1.6 | 85680 | 21 | AA722299 | BAC containing rep |
| 26 | 39.8 | 1.6 | 2832 | 23 | ABL10537 | Drosophila melanog |
| 27 | 39.8 | 1.6 | 5438 | 23 | ABL10536 | Drosophila melanog |
| 28 | 39.8 | 1.6 | 6378 | 24 | ABQ67027 | Human angiogenesis |
| 29 | 39.8 | 1.6 | 6378 | 24 | ABL32176 | Human immune syste |
| 30 | 39.8 | 1.6 | 136284 | 24 | ABK83575 | Human cDNA differe |
| 31 | 39.6 | 1.5 | 2215 | 24 | AB199688 | Mouse ischaemic co |
| 32 | 39.6 | 1.5 | 2375 | 21 | AA449881 | Human pancreatic p |
| 33 | 39.6 | 1.5 | 10266 | 17 | AA733007 | Mouse SKY-related |
| 34 | 39.6 | 1.5 | 14704 | 13 | AAQ20685 | PKS 741 insert con |
| 35 | 39.2 | 1.5 | 1935 | 23 | ABL27239 | Drosophila melanog |
| 36 | 39.2 | 1.5 | 3263 | 23 | ABL10403 | Drosophila melanog |
| 37 | 39.2 | 1.5 | 4374 | 23 | ABL27238 | Drosophila melanog |
| 38 | 39.2 | 1.5 | 16962 | 23 | ABL10402 | Drosophila melanog |
| 39 | 39 | 1.5 | 6354 | 24 | ABL32588 | Human immune syste |
| 40 | 38.8 | 1.5 | 4590 | 22 | AAH24065 | Yeast AOD9604-asso |
| 41 | 38.8 | 1.5 | 4829 | 23 | ABL13399 | Drosophila melanog |
| 42 | 38.8 | 1.5 | 39746 | 23 | ABL13398 | Drosophila melanog |
| 43 | 38.6 | 1.5 | 5101 | 20 | AA13453 | Drosophila melanog |
| 44 | 37.8 | 1.5 | 1037 | 21 | AA59242 | Enterococcus faeca |
| 45 | 37.8 | 1.5 | 1159 | 21 | AA59240 | Exons E, C and A o |
| | | | | | | An EcoRI fragment |

ALIGNMENTS

RESULT 1
AAV02872
ID AAV02872 standard; cDNA; 2556 BP.

XX AC AAV02872;

XX DT 28-APR-1998 (first entry)

XX XX Human hTAFII105 cDNA.

XX TATA-binding protein associated binding factor 105; human; activator;
KW hTAFII105; transcription factor; TFIID; transcriptional activation;
KW antibodies; diagnosis; therapy; biopharmaceutical industry; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1...2406

XX FT /*tag= a

XX FT /product= hTAFII105

XX FT /note= "human tata binding protein associated
factor 105 partial sequence"

XX PN US5710025-A.

XX PD 20-JAN-1998.

XX XX 02-OCT-1996; 96US-0725012.

XX PF 02-OCT-1996; 96US-0725012.

XX PR 02-OCT-1996; 96US-0725012.

XX XX (REGC) UNIV CALIFORNIA.

XX PA

```
Db 1739 TGGTTGGCACACTCATTGAGTCAATGATGAACATTTCTTTTATTGGAGCTCTAC 1798
Qy 1801 AAAAGAGAACTTTAGACATTTGGTAAAAAGCATGACATTTACAGAACTTTAACTCTGATGCTG 1860
Db 1799 AAAAGAGAACTTTAGACATTTGGTAAAAAGCATGACATTTACAGAACTTTAACTCTGATGCTG 1858
Qy 1861 TGAACCTTGATCTCCCAAGCACAGGAACTACAGGAGCTTCTAGAAAACTGACTG 1920
Db 1859 TGAACCTTGATCTCCCAAGCACAGGAACTACAGGAGCTTCTAGAAAACTGACTG 1918
Qy 1921 CAATTGCTCAGCATCGAATGACTTACTTACAAGCAAGTGAATAATTTACATCTGCTGCTAGT 1980
Db 1919 CAATTGCTCAGCATCGAATGACTTACTTACAAGCAAGTGAATAATTTACATCTGCTGCTAGT 1978
Qy 1981 ATACCAAGTCAAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAGCAGAGAAAGG 2040
Db 1979 ATACCAAGTCAAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAGCAGAGAAAGG 2038
Qy 2041 ATTTGGAGAGAGAAATGTTTACTTAAGGCAAGCAAGAGCTGCTTCTAATAAGAGATC 2100
Db 2039 ATTTGGAGAGAGAAATGTTTACTTAAGGCAAGCAAGAGCTGCTTCTAATAAGAGATC 2098
Qy 2101 CAGAACAGCTCAGATTTAAAGCAGAGAAAGCCAAAGAGTTACAGCAATTTGGAATTTGCACAGA 2160
Db 2099 CAGAACAGCTCAGATTTAAAGCAGAGAAAGCCAAAGAGTTACAGCAATTTGGAATTTGCACAGA 2158
Qy 2161 TACAGCATAGAGAGCTAATCTCAGAGCTTTTGAGCTATTTGGACCAAGAGAGAGAC 2220
Db 2159 TACAGCATAGAGAGCTAATCTCAGAGCTTTTGAGCTATTTGGACCAAGAGAGAGAC 2218
Qy 2221 CACTAGATCTGGAAATTCAGAGCTTAAAGACACACCTTCTTGGCTTCGGACATCCAGCC 2280
Db 2219 CACTAGATCTGGAAATTCAGAGCTTAAAGACACACCTTCTTGGCTTCGGACATCCAGCC 2278
Qy 2281 TGACAGCCACCAACAGCTGATCTGCTCAAGAAATCACAGAAATCTGCTCAGGAGCTTGA 2340
Db 2279 TGACAGCCACCAACAGCTGATCTGCTCAAGAAATCACAGAAATCTGCTCAGGAGCTTGA 2338
Qy 2341 TATTTTGTATGGAACAGAGAGGAGATGAAGTATTCGAGCTCTATACCTGCGCCCTTC 2400
Db 2339 TATTTTGTATGGAACAGAGAGGAGATGAAGTATTCGAGCTCTATACCTGCGCCCTTC 2398
Qy 2401 TGAAGTGACCACTCCACTCTTCCATCCACATCTTGTATTTACTGCGCAAGAGACACA 2460
Db 2399 TGAAGTGACCACTCCACTCTTCCATCCACATCTTGTATTTACTGCGCAAGAGACACA 2458
Qy 2461 AAGCATTTGTCAGCTGCTGCTGAAATTTCAATTTCTGGAAAAATAACACCAACATGAAAGAG 2520
Db 2459 AAGCATTTGTCAGCTGCTGCTGAAATTTCAATTTCTGGAAAAATAACACCAACATGAAAGAG 2518
Qy 2521 CATTTGTTTACGATTAGAACTTTTATTAACCTTTACCTAT 2558
Db 2519 CATTTGTTTACGATTAGAACTTTTATTAACCTTTACCTAT 2556
```

RESULT 2

AZ90465
ID AZ90465 standard; DNA: 2556 BP.

XX AC AZ90465;

XX DT 06-JUN-2000 (first entry)

XX DE Transcription factor subunit TAFII105 polypeptide encoding DNA.

XX KW TATA box-binding protein associated factor II 105; TAFII105; cancer;
KW transcription factor; apoptosis; cytostatic; immunosuppressive;
KW antiinflammatory; virucide; antibacterial; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

```
FT CDS 1..2556  
FT /*tag= a  
FT /transl_except= (pos: 2404..2406, aa: Glx)  
FT /transl_except= (pos: 2499..2501, aa: Glx)  
FT /note= "the stop codon is not indicated"  
PN WO200012699-A1.  
XX 09-MAR-2000.  
XX 25-AUG-1999; 99WO-IL00464.  
XX 27-AUG-1998; 98IL-0125971.  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX Dikstein R, Yamit-hezi A;  
XX WPI: 2000-256640/22.  
XX P-PSDB: AAY57279.  
XX Polypeptide encoding TATA box binding protein associated factor II 105  
XX useful for treating e.g. cancers and inducing apoptosis has a dominant  
XX negative effect on the normal biological activity of the binding  
XX protein -  
XX Claim 12; Fig 1; 48pp; English.  
XX This DNA encodes a polypeptide comprising a (modified) fragment (I) of  
XX a TATA box-binding protein associated factor II 105 (TAFII105). A  
XX pharmaceutical composition comprising (I) or the polynucleotide or an  
XX inhibitor or antagonist of (I) is useful for treating cancers and  
XX inducing apoptosis in pathological cells. The composition is also useful  
XX for treating autoimmune diseases, inflammatory processes and viral or  
XX bacterial infections.  
SQ Sequence 2556 BP; 797 A; 614 C; 514 G; 631 T; 0 other;
```

Query Match 99.1%; Score 2534; DB 21; Length 2556;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2556; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 GGGACCCCTGGTGACCAAAAGTGGCTCCGGTCAGCGGCCCTCTCTAAAGTCAGCAGCGGCCCT 60
Db 1 GGGACCCCTGGTGACCAAAAGTGGCTCCGGTCAGCGGCCCTCTCTAAAGTCAGCAGCGGCCCT 60

Qy 61 AGGCTGCCTGCTCCTCAGATAGTCGCCGTGAAGGCCCCCAACACCCAGCAATCCAGTTT 120
Db 61 AGGCTGCCTGCTCCTCAGATAGTCGCCGTGAAGGCCCCCAACACCCAGCAATCCAGTTT 120

Qy 121 CTGCTAAATTTGCAGCTTCTCCAGGAACCGTTTTGATTTAAAGTAACAGTGGTCCGTTG 180
Db 121 CTGCTAAATTTGCAGCTTCTCCAGGAACCGTTTTGATTTAAAGTAACAGTGGTCCGTTG 180

Qy 181 ATGTTGGTATCTCTCAGCAAAACTGTAAAGAGCCGAGACCAAGTAACATAACCTCA 240
Db 181 ATGTTGGTATCTCTCAGCAAAACTGTAAAGAGCCGAGACCAAGTAACATAACCTCA 240

Qy 241 AGGCCAGCAGTACCAGCAATCTCTCAACAGTCAAAATCTGTACAGTCCGCACTTAGC 300
Db 241 AGGCCAGCAGTACCAGCAATCTCTCAACAGTCAAAATCTGTACAGTCCGCACTTAGC 300

Qy 301 TCACAAATTAATCAAGAAAGTGGCAGTGACACCTGTTAAAAAATTTGGCAAAATAGGAAC 360
Db 301 TCACAAATTAATCAAGAAAGTGGCAGTGACACCTGTTAAAAAATTTGGCAAAATAGGAAC 360

Qy 361 ACTGTGGTAACCACTGTTCCGAGGCTTCCCTCAGTACAACTGTGGCTGTGCCAACCAGT 420
Db 361 ACTGTGGTAACCACTGTTCCGAGGCTTCCCTCAGTACAACTGTGGCTGTGCCAACCAGT 420

Qy 421 RGTGCTACAGTTACTCTCTGGAAGCCATTTGAATCTGTAACTACCTGTAACCTTCAAG 480
Db 421 -GTGCTACAGTTACTCTCTGGAAGCCATTTGAATCTGTAACTACCTGTAACCTTCAAG 479

QY 481 TTTGGAGCATCAACACCTCCCTTCAATGAGCCCAATCTTAAGCAGAGAACTCAGCAGC 540
Db TTTGGAGCATCAACACCTCCCTTCAATGAGCCCAATCTTAAGCAGAGAACTCAGCAGC 539
QY 541 TGTTCAGATTAATCTTTTCGCGCAATGCTAGAAATGTGAAGAAATGCAAGAACTTCCT 600
Db TGTTCAGATTAATCTTTTCGCGCAATGCTAGAAATGTGAAGAAATGCAAGAACTTCCT 599
QY 601 TGCATTTGTTAAATAACTAGCATGTAGTGATCACAGTCCCTGAAATGGGCAAAATGT 660
Db TGCATTTGTTAAATAACTAGCATGTAGTGATCACAGTCCCTGAAATGGGCAAAATGT 659
QY 661 GAAGAACTGGTGGAAACAATTTTGGATGCAAAATCGAAGCAGAGAAATTTACTAGGAA 720
Db GAAGAACTGGTGGAAACAATTTTGGATGCAAAATCGAAGCAGAGAAATTTACTAGGAA 719
QY 721 ACTGTATGTGAACCTCAAGTCTTCACCTCAGGCTCAGCTGAGTTCCTTTCTTAAGAAAAG 780
Db ACTGTATGTGAACCTCAAGTCTTCACCTCAGGCTCAGCTGAGTTCCTTTCTTAAGAAAAG 779
QY 781 CGTGGTTCCTTACGACAACTTCTGCCTTAACCTCCAGAGCTTCATCCAGCAATGTGTCA 840
Db CGTGGTTCCTTACGACAACTTCTGCCTTAACCTCCAGAGCTTCATCCAGCAATGTGTCA 839
QY 841 GCAGACTTCTAGTGACATGGTCAATGTCTACCTGCTACTACACAGTAAACAATTCCTCTGT 900
Db GCAGACTTCTAGTGACATGGTCAATGTCTACCTGCTACTACACAGTAAACAATTCCTCTGT 899
QY 901 GGTGCAACTACAGTTCCTCAGCCAGTCTGAAAGTCAATATTTGTTCTTGGAGCAAC 960
Db GGTGCAACTACAGTTCCTCAGCCAGTCTGAAAGTCAATATTTGTTCTTGGAGCAAC 959
QY 961 AGCACCAGAACTGTGTAGTGCACAACTTTGAACCCACTTGTGTCCAGTGGGAGCAAA 1020
Db AGCACCAGAACTGTGTAGTGCACAACTTTGAACCCACTTGTGTCCAGTGGGAGCAAA 1019
QY 1021 AGCTGAGTTGTGACACTTTCATCTGTGGGCCCACTGTGCAACAGGAGGAAACACAGC 1080
Db AGCTGAGTTGTGACACTTTCATCTGTGGGCCCACTGTGCAACAGGAGGAAACACAGC 1079
QY 1081 TGGAACTGCTTTCACACTTCAAAACCACTGTGACATCTGTGGCAACACAGTGCAC 1140
Db TGGAACTGCTTTCACACTTCAAAACCACTGTGACATCTGTGGCAACACAGTGCAC 1139
QY 1141 CACGGTCTCAGTGCACCTGAAAGCCAGTGTCTGTGGAACAGAGTAAACACTGTCCCT 1200
Db CACGGTCTCAGTGCACCTGAAAGCCAGTGTCTGTGGAACAGAGTAAACACTGTCCCT 1199
QY 1201 TCCAGCAGTAACCTTTTGGAGAACTTCAGGTGACGATATTTGTCTTCCATCTGTGAAACC 1260
Db TCCAGCAGTAACCTTTTGGAGAACTTCAGGTGACGATATTTGTCTTCCATCTGTGAAACC 1259
QY 1261 TGTGTTCTTCTCTGCTGGGACACATCTGCAAGCTGTATTGGGACTCCAGTCCAAAT 1320
Db TGTGTTCTTCTCTGCTGGGACACATCTGCAAGCTGTATTGGGACTCCAGTCCAAAT 1319
QY 1321 CAAACTTGCCCGCGGCCCTGTCTTTTCAACACAGCTGGGATTCACACAGGAGTTC 1380
Db CAAACTTGCCCGCGGCCCTGTCTTTTCAACACAGCTGGGATTCACACAGGAGTTC 1379
QY 1381 AAGCAAGCAACTATTCCTATTTGTTTCACTAGTTCAGCAGCTTCAGGAGGCAATGAAA 1440
Db AAGCAAGCAACTATTCCTATTTGTTTCACTAGTTCAGCAGCTTCAGGAGGCAATGAAA 1439
QY 1441 ACAAGTGACCAAAATTCACATTCCTCAACATTTGACATTCAGAAATGTGGACAGAGAC 1500
Db ACAAGTGACCAAAATTCACATTCCTCAACATTTGACATTCAGAAATGTGGACAGAGAC 1499
QY 1501 GATGCCAGTGAACACCATTAATACCTACTAGTCAGTTCCTCCAGCTTCCATTTAAAGCA 1560
Db GATGCCAGTGAACACCATTAATACCTACTAGTCAGTTCCTCCAGCTTCCATTTAAAGCA 1559

QY 1561 AATTACCTCTGCTCGGAAATAAAATTTCTGTCACTTCAAGCATCTCTACTCTCAGAAAAATA 1620
Db AATTA-CTCTGCTCGGAAATAAAATTTCTGTCACTTCAAGCATCTCTACTCTCAGAAAAATA 1618
QY 1621 GAATAAAGAGAATGTAAACATCATGCTTCGAGATGAGGATGACATCAATGTGTGACTT 1680
Db GAATAAAGAGAATGTAAACATCATGCTTCGAGATGAGGATGACATCAATGTGTGACTT 1678
QY 1681 CTATGTCAGGGTCAACCTTAATGAAGAAATGCCTGCATCTTAGCAACAACTCTGAAT 1740
Db CTATGTCAGGGTCAACCTTAATGAAGAAATGCCTGCATCTTAGCAACAACTCTGAAT 1738
QY 1741 TGGTTGGCACACTCATTCAGTCAATGAAGATGAACCAATTTCTTTTATTTGGAGCTCTAC 1800
Db TGGTTGGCACACTCATTCAGTCAATGAAGATGAACCAATTTCTTTTATTTGGAGCTCTAC 1798
QY 1801 AAAGAGAATCTTAGACATTTGGTAAAGCATGACATTTACAGAACTTAATCTGTAGTGTG 1860
Db AAAGAGAATCTTAGACATTTGGTAAAGCATGACATTTACAGAACTTAATCTGTAGTGTG 1858
QY 1861 TGAACCTTGATCTCCCAAGCAACACAGAGACTACGAGGCTTCTAGAAAACTGACTG 1920
Db TGAACCTTGATCTCCCAAGCAACACAGAGACTACGAGGCTTCTAGAAAACTGACTG 1918
QY 1921 CAATTCCTCAGCATCGAATGACTTACTACAGGCAAGTGAATAATACATCTGTGTAGTG 1980
Db CAATTCCTCAGCATCGAATGACTTACTACAGGCAAGTGAATAATACATCTGTGTAGTG 1978
QY 1981 ATACAGGTCACAGCTCAAAATTTCTTGAAGCTGGATCAATTTGGAGAGCAGAGAAAG 2040
Db ATACAGGTCACAGCTCAAAATTTCTTGAAGCTGGATCAATTTGGAGAGCAGAGAAAG 2038
QY 2041 ATTTGGAAGAAAGAGAAATGTTTAAAGCAGCCAGAGTCTGTCTTAATAAAGAGATC 2100
Db ATTTGGAAGAAAGAGAAATGTTTAAAGCAGCCAGAGTCTGTCTTAATAAAGAGATC 2098
QY 2101 CAGAACAGCTGAGATTAAGCAGAGAAAGCAGAGTTCAGCAATTTGGAATTTGCACAGA 2160
Db CAGAACAGCTGAGATTAAGCAGAGAAAGCAGAGTTCAGCAATTTGGAATTTGCACAGA 2158
QY 2161 TACAGCATAGAGAGCTTAATCTCAGAGCTTTCAGCTATTGGACCAGAGAGAGAC 2220
Db TACAGCATAGAGAGCTTAATCTCAGAGCTTTCAGCTATTGGACCAGAGAGAGAC 2218
QY 2221 CACTAGAATCTGGAATTTGAGGGCTTAAAGACAACTTCTTGTCTTCTGGACATCCAGCC 2280
Db CACTAGAATCTGGAATTTGAGGGCTTAAAGACAACTTCTTGTCTTCTGGACATCCAGCC 2278
QY 2281 TGACAGCCACCAACAGTTCGATCGTCCAAAGATCAGGAAATCTGCCTCAGGAGACTTGA 2340
Db TGACAGCCACCAACAGTTCGATCGTCCAAAGATCAGGAAATCTGCCTCAGGAGACTTGA 2338
QY 2341 TATTTTGTATGGAACAGGAGGAGATGAAGTATTCTCGAGCTTATACCTGGCCCTTC 2400
Db TATTTTGTATGGAACAGGAGGAGATGAAGTATTCTCGAGCTTATACCTGGCCCTTC 2398
QY 2401 TGAAGTGACACTCCACTCTTCCATCCACATCTTGTCTATTATTCGCAAGAGACACA 2460
Db TGAAGTGACACTCCACTCTTCCATCCACATCTTGTCTATTATTCGCAAGAGACACA 2458
QY 2461 AAGCATTTGTGACTGTCTGAAATTTCAATTTCTGGAAAAATACACCAATGAAAGAG 2520
Db AAGCATTTGTGACTGTCTGAAATTTCAATTTCTGGAAAAATACACCAATGAAAGAG 2518
QY 2521 CATTTGTTACGATTAAGAACTTTTAACTCTTACCTAT 2558
Db CATTTGTTACGATTAAGAACTTTTAACTCTTACCTAT 2556

RESULT 3
AAS73655
ID AAS73655 standard; cDNA; 3468 BP.
XX

QY 1261 TGTGTTTCTCTGCTGGGACCACATCTG-CAAGCCTGTTATTGGGACTCCAGTTCAAA 1319
Db 1260 TTTGTTTCTCTGCTGGGACCACATCTGACAAGCCTGTTATTGGGACTCCAGTTCAAA 1319
QY 1320 TCAAACTTCCCGACGGGCCCTGTCTTTTCAACACGACCTGGGATTCCAAACAGGCAGTT 1379
Db 1320 TCAAACTTCCCGACGGGCCCTGTCTTTTCAACACGACCTGGGATTCCAAACAGGCAGTT 1379
QY 1380 CAAG-CAAGCAACTATTCTCATTGTTTTCACGTAGTTTCAGCAGCCTTCAGGAGGCAATGAA 1438
Db 1380 CAAGTCAAGCAACTATTCTCATTGTTTTCACGTAGTTTCAGCAGCCTTCAGGAGGCAATGAA 1439
QY 1439 AAACAAGTCAACCAATTTTCAATCTCCTCAACATTGACCATTCAGAAATGTGGACAGAG 1498
Db 1440 AAACAAGTCAACCAATTTTCAATCTCCTCAACATTGACCATTCAGAAATGTGGACAGAG 1499
QY 1499 ACGATGCCAGTGAACACCAATAACTACTAGTCAAGTTTCCCTCCAGCTTCCATTTCTAAG 1558
Db 1500 ACGATGCCAGTGAACACCAATAACTACTAGTCAAGTTTCCCTCCAGCTTCCATTTCTAAG 1559
QY 1559 CAANTTACCTCTGCTGGAAATAAATCTGTCACTTCAAGCATCTCTTACTCAGAAAAA 1618
Db 1560 CAATTA-CTCTGCTGGAAATAAATCTGTCACTTCAAGCATCTCTTACTCAGAAAAA 1618
QY 1619 TAGAATAAAGAGATGTAACATCATGCTTCCGAGATGAGGATGACATCAATGATGTGAC 1678
Db 1619 TAGAATAAAGAGATGTAACATCATGCTTCCGAGATGAGGATGACATCAATGATGTGAC 1678
QY 1679 TTCTATGGCAGGGTCAACCTTAAGTAGAATAAGCCTGCATCTTAGCAACAACCTCTGA 1738
Db 1679 TTCTATGGCAGGGTCAACCTTAAGTAGAATAAGCCTGCATCTTAGCAACAACCTCTGA 1738
QY 1739 ATTGGTTGGCACACATTCAGTCATGTAAGATGAACCATTTCTTTTATTGGAGCTCT 1798
Db 1739 ATTGGTTGGCACACATTCAGTCATGTAAGATGAACCATTTCTTTTATTGGAGCTCT 1798
QY 1799 ACAAAAGAGAACTTTAGACATTTGGTAAAGCATGACATTCAGAACATTAACCTCTGATGC 1858
Db 1799 ACAAAAGAGAACTTTAGACATTTGGTAAAGCATGACATTCAGAACATTAACCTCTGATGC 1858
QY 1859 TGTGAACCTTGATCTCCCAAGCAACACAGGACGACTACGAGGCCCTTAGAAGAAACGAC 1918
Db 1859 TGTGAACCTTGATCTCCCAAGCAACACAGGACGACTACGAGGCCCTTAGAAGAAACGAC 1918
QY 1919 TGAATTTGCTCAGCATCGAATGACTACTTACAAGGCAAGTGAATAATACATCTGTGTAG 1978
Db 1919 TGAATTTGCTCAGCATCGAATGACTACTTACAAGGCAAGTGAATAATACATCTGTGTAG 1978
QY 1979 TGATACCAGGTCACAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAACAGAAA 2038
Db 1979 TGATACCAGGTCACAGCTCAAAATTTCTTGAAGAGCTGGATCAATTTGGAGAACAGAAA 2038
QY 2039 GGATTTGGAAGAGAGAAATGTTACTTAAAGCAGCAAGAGTCTGCTTAATAAAGAGA 2098
Db 2039 GGATTTGGAAGAGAGAAATGTTACTTAAAGCAGCAAGAGTCTGCTTAATAAAGAGA 2098
QY 2099 TCCAGACAGCTGAGATTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGGAATTCGACA 2158
Db 2099 TCCAGACAGCTGAGATTAAAGCAGAAAGCCAAAGAGTTACAGCAATTTGGAATTCGACA 2158
QY 2159 GATACAGCATAGACAGCTTAATCTCACAGCTCTTCAGCTATTGGACCAAGGAAGAAG 2218
Db 2159 GATACAGCATAGACAGCTTAATCTCACAGCTCTTCAGCTATTGGACCAAGGAAGAAG 2218
QY 2219 ACCACTAGAACTGGAATTTGAGGCTTAAAGACAACTCTTCTGCTTCTGGACATCCAG 2278
Db 2219 ACCACTAGAACTGGAATTTGAGGCTTAAAGACAACTCTTCTGCTTCTGGACATCCAG 2278
QY 2279 CCGTACAGCCCAACAGTTGCTGCTCCCAAGATTCAGGAATCTGCCTCAGGACTT 2338
Db 2279 CCGTACAGCCCAACAGTTGCTGCTCCCAAGATTCAGGAATCTGCCTCAGGACTT 2338
QY 2339 GATATTTTGTATGGAACAGGACGGAGNTGAAGTATTCTCCAGCTCTATACCTGGCCCT 2398

Db 2339 GATATTTTGTATGGAACAGGACGGAGATGAAGTATTCTCGAGCTCTATACCTGGCCCT 2398
QY 2399 TCTGAAGTCAACCACTCCACTCTTCCATCCACATCCTTGTCTATTACTGCGCAAGAGACA 2458
Db 2399 TCTGAAGTCAACCACTCCACTCTTCCATCCACATCCTTGTCTATTACTGCGCAAGAGACA 2458
QY 2459 CAAAGCATTTGTCACCTGCTGAAATTTCAATTTCTGAAAAATAA-CACCAACATGAAA 2517
Db 2459 CAAAGCATTTGTCACCTGCTGAAATTTCAATTTCTGAAAAATAATCACCACATGAAT 2518
QY 2518 GAGCATTTGTTAGGATTAG-AACTTTATTAACTCTTACCTAT 2558
Db 2519 GAGCATTTGTTAGGATTAGAAACTTTATTAACTCTTACCTAT 2560

RESULT 4

AAAT42217
ID AAAT42217 standard; cDNA; 3603 BP.
XX
AC AAAT42217;
XX
DT 27-JAN-1997 (first entry)
XX
DE Human TATA-binding protein associated factor hTAFII130 gene.
KW Human; TATA-binding protein; TBP associated factor; TFIID; holoenzyme;
KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
KW lambda-gt11; expression library; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2214
FT /tag= a
FT /product= human TAFII130
XX
PN US5534410-A.
XX
PD 09-JUL-1996.
XX
PF 28-JAN-1993; 93US-0013412.
XX
PR 28-JAN-1994; 94US-0188582.
PR 28-JAN-1993; 93US-0013412.
PR 30-JUN-1993; 93US-0087119.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI Tjian R, Wang E, Weinzierl ROJ;
XX
DR WPI; 1996-333245/33.
DR P-PSDB; AAW06084.
XX

Screen for cpds. that bind human TATA-binding protein associated factor - by testing ability to bind to polypeptide fragments of the factor, useful as (ant)agonists of transcription factors involved in disease.

Examples; Column 105-112; 86pp; English.

This is the nucleotide sequence encoding the human TATA-binding protein (TBP) associated factor (TAF) designated TAFII130. The protein is a component of the TFIID fraction required for reconstituting RNA polymerase II in vitro transcription activity. The encoded protein has an estimated mol. wt. of 130 kD by SDS-PAGE.

The invention relates to purified proteins involved in transcription by RNA polymerase II, the RNA polymerase which transcribes messenger RNA. RNA polymerase II transcription proceeds in vitro upon addition of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J to RNA polymerase II holoenzyme. Fraction TFIID has been shown to contain a TBP and other TAFs. Purification of TFIID and separation of

CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIID fraction allowed cloning of the corresp.
 CC genes from lambda-gt11 expression libraries.

XX
 SQ Sequence 3603 BP; 969 A; 961 C; 830 G; 843 T; 0 other;
 Query Match 13.2%; Score 338.2; DB 17; Length 3603;
 Best Local Similarity 63.9%; Pred. No. 1.6e-84;
 Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;
 QY 1586 TCTGTCATCTCAAGCATCTCTACTCAGAAAAATAGATAAAGAGAAATGTAAACATCATG 1645
 DB 1377 TGTCTCGCACACAGCAGCTGCTGCACAGAAAAATAAACTCAAGGAGCTGGGGAGGTTG 1436
 QY 1646 CTTCCGAGATGAGGATGACATCAATGATGTGACTTCTATGCGAGGGGTCAACCTTAATGA 1705
 DB 1437 GTTTCGGGACGATGACATTAATGATGTGCTGATGCTGCTGAGTAAACTTGTGAGA 1496
 QY 1706 AGAAATGCCCTGCTTAGCAACAACTCTGAATTTGGTGGCACATCTTCAGTCTATG 1765
 DB 1497 AGAAGTGCAGAAATATTAGCCAGCACTCTGAATTTGGTGGCACCTTAACGCGGCTCTG 1556
 QY 1766 TAAAGATCAACCACTTCTCTTTTATTGGAGCTCTACAAAAGAGAAATCTTAGACATTTGGTAA 1825
 DB 1557 TAAAGATGAACCTTCTCTCCCAAGCGCTTTCGACAGAGAAATATTAGAAATAGGTAA 1616
 QY 1826 AAGCATGACATTCAGAACTTAACCTCTGATGCTGTGAATTTGATCTCCCAAGCAACACA 1885
 DB 1617 AAAACATGGTAAACGGAATACATCCAGATGTAGTAACTATGATCATCATGCCAGCA 1676
 QY 1886 GGAACGACTAGGAGCGCTCTAGAAAACCTGACTGCAATTTGCTCAGCATCGAATGACTAC 1945
 DB 1677 ACAAGCTACAGAACTCTGTAGAGAAATATCAGAACAGCTCAGCAGAGAACTTTTC 1736
 QY 1946 TTACAAGCAGTGAATTAACATCTCTGTGTAGTATACCAAGTCTCAGCTCAAAATTTCT 2005
 DB 1737 TTACAAGTATGACGACAGATATGAGCGGAGTGACGTCCGGSCACAGCTCAAGTTTT 1796
 QY 2006 TGAAGACTGTGATCAATTTGGAGAGCAGAGAAAGGATTTGGAAGAAAGAGAAATGTTACT 2065
 DB 1797 TGAACAGCTTTGATCAAAATCCAAAAGCAGAGAGAGATGAGCAGGCGGAGATCTCTGAT 1856
 QY 2066 TAAGCAGCAGCAAGATGCTCTTAATAAAGAGATCCAGAACAGCTGAGATTAAAGCAGAA 2125
 DB 1857 GAGGCGAGCAAAAGTCTCGGTCAAGACAGAAAGATCCAGACAGTTAAGGCTGAAACAGAA 1916
 QY 2126 AGCCAAAGAGTTACAGCAATTTGGAACCTTGACAGATACAGATAGAGACGCTAATCTCAC 2185
 DB 1917 GGCAAGAGAGATGACGACACAGGAATGGCACAAATGAGACAGCGGCGGACCACTCAC 1976
 QY 2186 AGCTTTGACAGCTATTGGACCAAGGAAGAGAGACCACTAGA-----ATC 2230
 DB 1977 AGCACTAGCAGCATCGGGCCCGGACGAGAAAAAGAGAAAGTGGACTGTCCGGGCGGGGCTC 2036
 QY 2231 TGGAAATGAGGGCTTAAAGACACCTTCTGTCTTCTGGGACATCCAGCTGACAGCCAC 2290
 DB 2037 AGGAGCAGAGGGTCTGGGCGCCGCTCAGTGTGTCCTCCAGGAGCTGCGGGTGTGCGGAACCCC 2096
 QY 2291 CAACAGCTTGCATCGTCCAGAAATCAGAGAAATCTGCTCAGGACCTTGATATTTTGTAT 2350
 DB 2097 CAGACAGTTTCCCGGACAAAGATCAGCGGGTCAACCTCAGGACCTCATATTTTGT 2156
 QY 2351 GGAACAGAGAGGGAGATGAGTATTTCTCGAGCTCTATACCTGGCCCTTCTGGAAGTAC 2409
 DB 2157 AGAAATGAACGTGAGACAGCCATTCACCTGCTCTACAAAGCAATTCCTTAAAGTGAC 2215

RESULT 5

AAT79595

ID AAT79595 standard; cdna; 3603 BP.

XX

AC AAT79595;

XX

DT 08-OCT-1997 (first entry)
 DE TATA-binding protein associated factor, hTAFII130 cDNA.
 XX
 KW TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP;
 KW Initiation; ds.
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2214
 FT /*tag= a
 FT /note= "no start codon"

FT US5637686-A.

PN 10-JUN-1997.

PD 28-JAN-1993; 93US-0013412.

PF 28-JAN-1994; 94US-0188582.

XX 28-JAN-1993; 93US-0013412.

PR 30-JUN-1993; 93US-0087119.

PR 09-MAY-1996; 96US-0646715.

XX (REGC) UNIV CALIFORNIA.

PA Comai L, Dynolact BD, Hoey T, Ruppert S, Tanese N;

PI Tjian R, Wang E, Weinzierl ROJ;

XX WPI; 1997-319113/29.

DR P-PSDB; AAW25019.

XX Nucleic acids encoding human TATA-binding protein associated factor

XX (TAF) peptide(s) - for production of recombinant peptide(s), used

XX for modulating transcription of TAFs

PS Claim 11; Column 105-112; 86pp; English.

XX AAT79595 encodes human TATA-binding protein associated factor (TAF)

XX polypeptide, hTAFII130 (mol. weight 130kD). TAF peptides derived

XX from hTAFII130 alpha, hTAFII130 beta, hTAFII140, hTAFII70, hTAFII100,

XX hTAFII130, hTAFII250, hTAFII48 and hTAFII10 and nucleic acids encoding

XX them, are used to modulate transcription, including transcription

XX initiation. TAFs are nuclear proteins involved in RNA polymerase I,

XX II and III transcription. The peptides act by binding to a different

XX TAF, an activator, or TBP (TATA-binding protein) or competitively

XX inhibiting association of a TAF domain with another compound, typically

XX a protein like TBP or another TAF, an activator, or DNA.

XX Sequence 3603 BP; 969 A; 961 C; 830 G; 843 T; 0 other;

SQ Query Match 13.2%; Score 338.2; DB 18; Length 3603;

Best Local Similarity 63.9%; Pred. No. 1.6e-84;

Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

QY 1586 TCTGTCATCTCAAGCATCTCTACTCAGAAAAATAGATAAAGAGAAATGTAAACATCATG 1645

DB 1377 TGTCTCGCACACAGCAGCTGCTGCACAGAAAAATAAACTCAAGGAGCTGGGGAGGTTG 1436

QY 1646 CTTCCGAGATGAGGATGACATCAATGATGTGACTTCTATGCGAGGGGTCAACCTTAATGA 1705

DB 1437 GTTTCGGGACGATGACATTAATGATGTGCTGATGCTGAGTAAACTTGTGAGA 1496

QY 1706 AGAAATGCCCTGCTTAGCAACAACTCTGAATTTGGTGGCACATCTTCAGTCTATG 1765

DB 1497 AGAAGTGCAGAAATATTAGCCAGCACTCTGAATTTGGTGGCACCTTAACGCGGCTCTG 1556

QY 1766 TAAAGATCAACCACTTCTCTTTTATTGGAGCTCTACAAAAGAGAAATCTTAGACATTTGGTAA 1825

DB 1557 TAAAGATGAACCTTCTCTCCCAAGCGCTTTCGACAGAGAAATATTAGAAATAGGTAA 1616

QY 1826 AAAGCATGACATTACAGAACTTAACCTCTGATGCTGTGAACCTTGATCTCCCAAGCAACACACA 1885
 DB 1617 AAAACATGATAACCAAGGAATATACATCAGATGTAGTAAGTTATATCATCATGCCACGCA 1676
 QY 1886 GGAACCACTACGAGGCTTCTAGAAAACCTGACTGCAATTTGCTCAGCATCGAATGACTAC 1945
 DB 1677 ACNAGGCTACAGATCTTGTAGANNATATCAGANACAGCTCAGCAGAAAGCACTTTTC 1736
 QY 1946 TTACAAGGCAAGTGAATATACATCTCTGTGTATGATACAGCTCAGCTCAAAATTTCT 2005
 DB 1737 TTACAAGGATGACGACAGATATGACGAGCGAGTGACCTCCGGGCACACCTCAAGTTT 1796
 QY 2006 TGAAGAGCTGGATCATTTGGAGAGCAGAGAAAGGATTTGGAGAAAGAAATGTTACT 2065
 DB 1797 TGAACAGCTTGATCAATCAAAATGAAAGCAGAGAGGATGACGAGGCGGAGATCTGAT 1856
 QY 2066 TAAGGAGCAGCAAGAGTCGTCTTAATAAGAAAGATCCAGAACAGCTCAGATTAAAGCAGAA 2125
 DB 1857 GAGGCGAGCAAGTCTCGGTCAAGACAGAGATCCAGAACAGTTAAGGCTGAAACAGAA 1916
 QY 2126 AGCCAAAGAGTTACAGCAATTTGGAACCTTGCACAGATACAGCATAGAGAGCGCTAATCTCAC 2185
 DB 1917 GCAAAAGGAGATGCAAGCAAGGAACTTGCACAAATAGACAGCGGAGCCCAACCTCAC 1976
 QY 2186 AGCTCTTGACGCTATTGGACCAAGGAAGAGAGACCACTAGA-----ATC 2230
 DB 1977 AGCACTAGCAGCGATCGGGCCCGAGGAAAGAAAGTGGACTGTCCGGGGCCGGGCTC 2036
 QY 2231 TGAATTTGAGGCTTAAAGACAACTTCTTCTTCTGGACATCCAGCTGACAGCCAC 2290
 DB 2037 AGGAGCAGAGGGTGGGGCCCGGTCTAGTGTCCAGGACCTCGGGTGGGAACCC 2096
 QY 2291 CAAACAGTTGCATGCTCCAAAGATCACGAGAACTCTCCCTCAGGAGCTTGATATTTTGTAT 2350
 DB 2097 CAGACAGTTCCGCGACGAAAGAAATCAGCGGGTCAACCTCAGGACCTCATATTTTGT 2156
 QY 2351 GGAACAGGAACGGGAGATGAACTATTCGAGCTCTATACCTGGGCTTCTGAAATGAC 2409
 DB 2157 AGAAAATGAACGTGAGACAAAGCATTCACCTGTCTACAAAGCATTCCTTAAAGTGAC 2215

RESULT 6

NAQ070731
 ID NAQ070731 standard; cDNA: 2214 BP.

XX
 AC
 XX
 XX
 DT
 XX
 DE

23-MAR-1995 (first entry)

DE TATA-binding protein-associated factor hTAFII130 cDNA.

XX
 KW TATA-binding protein associated factor; hTAFII130; ss; screening;
 KW diagnostic; therapeutic; gene transcription regulation.

OS Homo sapiens.

XX
 FH Key
 FT Location/Qualifiers
 FT CDS
 FT 1..2214
 FT /*tag= a

XX W09417087-A.

XX 04-AUG-1994.

XX 28-JAN-1994; 94WO-US01114.

XX 28-JAN-1993; 93US-0013412.

XX 30-JUN-1993; 93US-0087119.

XX (REGC) UNIV CALIFORNIA.

XX Comai L, Dynlact BD, Hoey T, Ruppert S, Tanese N;
 PI Tjlan R, Wang E, Weinzierl ROJ;

XX

DR WPI; 1994-264019/32.

XX P-PSDB; AAR56494.

XX TATA-binding protein associated protein factors - and

PT corresponding nucleotide sequence and deriv. antibodies, useful

PT in screening, diagnostics and therapeutics

XX Disclosure; Page 137-141; 180pp; English.

XX The TATA-binding protein associated factor hTAFII130 (including
 CC specific antibodies and fusion products) are used in drug screening,
 CC diagnostics and therapeutics. They are used in the development of
 CC specific biochemical assays for screening compounds that agonise or
 CC antagonise selected transcription factors involved in regulating
 CC gene expression associated with human pathology.

XX Sequence 2214 BP; 563 A; 691 C; 600 G; 360 T; 0 other;

Query Match

Best Local Similarity 13.2%; Score 337.2; DB 15; Length 2214;

Matches 535; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

QY 1586 TGTGTCACCTCAAGCATCTCTACTCAGAAAATAGAAAATGTAACATCATG 1645

DB 1377 TGTCTGGCACAAGCAGCTGCTGCACAGAAAATAAACAAGAGCCTGGGGAGGTTT 1436

QY 1646 CTTCCGAGATGAGGATGACATCAATCATCTGACTTCTATGGCAGGGTCAACCTTAATGA 1705

DB 1437 GTTTCGGGACGATGATGATTAATGATGTTGCATCGATGGTGGAGTAACTTGTGCA 1496

QY 1706 AGAAATGCTGCATCTTAGCAACAACCTCTGAATTTGGTGGCACACTCATTCAGTCATG 1765

DB 1497 AGAAGTGCAGAAATATATAGCCACGAACTCTGAATTTGGTGGCACCGCTAACCGGTCCTG 1556

QY 1766 TAAAGATGAACCATTTCTTTTATTTGGAGCTCTACAAAAGAGAAATTTAGACATTTGGTAA 1825

DB 1557 TAAAGATGAACCTTCTCTCCCAAGCGCTTTGCAGAGAAGAAATATTAGAAATAGGTAA 1616

QY 1826 AAAGCATGACATTACAGAACTTAACCTCTGATGCTGTGAACCTTGAATCTCCCAAGCAACACA 1885

DB 1617 AAACATGTTATACGGAATTAACATCCAGATGTAGTAAGTTATGTATCATCATGCCACGA 1676

QY 1886 GGAAGCATACGAGGCTTCTAGAAAACCTGACTGCAATTTGCTCAGCATCGAATGACTAC 1945

DB 1677 ACAAGGCTACAGAACTTTGTAGAGAAATATCAGAAAACAGCTCAGCAGAAAGCACTTTTC 1736

QY 1946 TTACAAGGCAAGTGAATATACATCTGTGTGTATGATACCAAGTCTCAGCTCAAAATTTCT 2005

DB 1737 TTACAAGGATGACGACAGATATGACGAGCGAGTGCAGTCCGGGCACAGCTCAAGTTT 1796

QY 2006 TCAAAAGCTTGGATCAATTTGGAGAAGCAGAGAAAGGATTTGGAGAAGAAAGAAATGTTACT 2065

DB 1797 TGAACAGCTTGATCAATTCGAAAAGCAGAGAGAGGATGATGACGAGGAGGAGAGCTGAT 1856

QY 2066 TAAGGAGCAGCAAGTGTCTTCTAATAAGAAAGATCCAGAACAGCTGAGATTTAAAGCAGAA 2125

DB 1857 GAGGCGAGCAAGTCTCGGTCAAGACAAGAAATCCAGAACAGTAAAGCTGNAACAGAA 1916

QY 2126 AGCCAAAGAGTTACAGCAATTTGGAACCTTGCACAGATACAGCATAGAGAGCGCTAATCTCAC 2185

DB 1917 GCAAAAGGAGATGCAAGCAAGGAACTTGCACAAATAGACAGCGGAGCCCAACCTCAC 1976

QY 2186 AGCTCTTGACGCTATTGGACCAAGGAAGAGAGACCACTAGA-----ATC 2230

DB 1977 AGCACTAGCAGCGATCGGGCCCGAGGAAAGAAAGTGGACTGTCCGGGGCCGGGCTC 2036

QY 2231 TGAATTTGAGGCTTAAAGACAACTTCTTCTTCTGGACATCCAGCTGACAGCCAC 2290

DB 2037 AGGAGCAGAGGGTTCGGGGCCCGGTCTAGTGTGCCAGCAGCTCGGGTGTGGAACCCC 2096

QY 2291 CAAACAGTTGCATCGTCCAGAAATCACGAGAAATCTGCCCTCAGGGACTTGTATTTTGTAT 2350

DB 2350

| | | | | |
|----------|---|---|--|------|
| D | b | 2097 | CAGACAGATTCACCCGCAAAAGAATCACGGGGGTCAACCTCAGGGACCTCATATTTTCTTTT | 2155 |
| Q | y | 2351 | GGAACAGCAAGCGGAGATGAAGTATTCTCGAGCTCTATATACCTGGCCCTTCTGGAAGTGA | 2408 |
| D | b | 2157 | AGAAATGAACGTGAGACAAGCCATTCACTGCTGCTCTACAAAGCATTCCTTAAGTGA | 2214 |
| RESULT 7 | | | | |
| A | B | L | O | 9423 |
| I | D | ABL09423 | standard; cDNA; 4520 BP. | |
| X | X | ABL09423; | | |
| X | X | 26-MAR-2002 | (first entry) | |
| X | X | Drosophila melanogaster | expressed polynucleotide SEQ ID NO 22751. | |
| K | W | Drosophila; | developmental biology; cell signalling; insecticide; | |
| K | W | pharmaceutical; gene; ss. | | |
| X | S | Drosophila melanogaster. | | |
| P | N | WO200171042-A2.. | | |
| P | D | 27-SEP-2001. | | |
| X | X | 23-MAR-2001; | 2001WO-US09231. | |
| X | X | 23-MAR-2000; | 2000US-191637P. | |
| P | R | 11-JUL-2000; | 2000US-0614150. | |
| X | P | (PEKE) | PE CORP NY. | |
| V | E | Venter JC, | Adams M, Li PWD, Myers EW; | |
| D | R | WPI; | 2001-656860/75. | |
| X | X | P-PSDB; | ABB65320. | |
| N | e | w | i | s |
| P | T | New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions - | | |
| X | X | Claim 1; | SEQ ID NO 22751; 2lpp + Sequence Listing; English. | |
| C | C | The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA sequences (ABLO1840-ABLI16175) and the encoded proteins (ABB57737-ABB72072). | | |
| C | C | The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. | | |
| X | X | Sequence 4520 BP; 1350 A; 1182 C; 1071 G; 917 T; 0 other; | | |
| Q | U | Query Match | 5.6%; Score 143.8; DB 23; Length 4520; | |
| M | a | Best Local Similarity | 50.8%; Pred. No. 1.4e-29; | |
| M | a | Matches 396; Conservative | 0; Mismatches 377; Indels 6; Gaps | |
| Q | y | 1653 | GATGAGGATGCATCAATCATGTGACTTCTATGGCAGGGGTCAACCTTAATGAAGAAAAT | 1712 |
| D | b | 2491 | GGTGACGATGATATCAACGATGTTCCGCCATGGGAGGTGTTAACTTGGCGGAGAGTCG | 2550 |
| Q | y | 1713 | GCCTGCATCTTAGCAAACAACTCTGAATTGGTTGGCACACATCATCTCACTGATGTAAGAAT | 1772 |
| D | b | 2551 | CAGCGAATTCCTGGCTGTGA---CCGAAACATCGCGCAGCATTCGATCCTCTGCAAAAGAT | 2607 |
| Q | y | 1773 | GAACCATTTCTTTTATTGGAGCTCTACAAAGAGAAATCTTAGACANTTGGTAAAAACAT | 1832 |
| D | b | 2608 | GAGTATTTCTCAATCTCCCTCGCTGCAAGCTAGAACTACGGGCAATTACTTCGGAGCGC | 2667 |

QY 1893 CTACGAGGCTTCTAGAAAACTGACTGCAATTTGCTGAGCATCGAATGACTACTTACAAG 1952
 DB 2785 CTGAAGAACATCGTTGAGAACTGGCTGTGATGCGGAGCACCGCATTTGATGTCTCATCAAG 2844
 QY 1953 GCAAGTGAATAATACATCTCTGTAGTAGATACACAGGTACACAGTCTCAAAATTTCTTGAAGAAG 2012
 DB 2845 TTGGATCCACGCTATGATGCCCGCCCAAGGATGTGCGGCTGAGATCAAGTTTCTCGAGGAG 2904
 QY 2013 CTGATCAATTTGGAGACGAGAGAAAGATTTGGAAGAAAGAGAAATGTTTAAAGGCA 2072
 DB 2905 CTGGACAAGGCCGAGCAGAGAGCAGAGGAACTGGAGCTGAGATGCTGCTGCGGGCA 2964
 QY 2073 GCCAGAGTCTTCTTAATAAGAGATCCAGAACAGCTGAGATTAAGCCAGAAAGCCAAA 2132
 DB 2965 GCCAAGTCACTGCTGAGGCTGGAAGATCCCGAGCAGGCAAGATGAAGGCGAGGCCCAAG 3024
 QY 2133 GAGTTACAGCAATTTGGAACTTGCACAGATACAGCATAGAGACGCTAATCTCACAGCTCTT 2192
 DB 3025 GAGATGCAACCGCGCGAATGGAGGAGTTGGTCAACGAGATGCCATCTGACGCGCTG 3084
 QY 2193 GCAGCTATTTGGACCAAGAGAGACCACTAGAAATTTGAGGGCTTAAAGAC 2252
 DB 3085 CAGCGATTTGGACCTCGGAAAAAGCTGAAGCTGGACGCGGAAACAGTCAGTTCTGGGAGCG 3144
 QY 2253 AACCTTCTGCTTCTGG---GACATCCAGCTGACAGCCCAACAGTTGCATCGTCCA 2309
 DB 3145 GGTTCAGTGGCGGCGAGTCTAAGCAGCTCGGGATCTGCGCGCAGCAGCTTACGGCCT 3204
 QY 2310 AGAATCAGGAAATCTGCTCAGGAGCTTGATATTTTGTATGGAACAGAGGAGGAGATG 2369
 DB 3205 CGCATAAACGCTGTGAACCTCGCGACATGCTCTTACATGGAGCAGAGCGGAGTTC 3264
 QY 2370 AAGTATCTCGAGCTATACCTGGCCCTTCTGAAGTGACACACTCCACTCTTCCATCCA 2428
 DB 3265 TGTGCGAGTTCCATGCTTCAAGACATACCTCAAGTATGCTGCTGCTGCTGCCCATCAA 3323

RESULT 10

ID AAT42210
 standard; cdna; 4615 BP.

XX AC AAT42210;

XX DT 27-JAN-1997 (first entry)

XX Drosophila TATA-binding protein associated factor dTAFII110 gene.

XX Drosophila; TATA-binding protein; TBP associated factor; TFIIID;

KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;

KW holoenzyme; lambda-gtl; expression library; ds.

XX Drosophila melanogaster.

XX Key Location/Qualifiers

XX CDS 538..3303

XX /*tag= a

XX /product= Drosophila TAFII110 protein

XX US5534410-A.

XX 09-JUL-1996.

XX 28-JAN-1993; 93US-0013412.

XX 28-JAN-1994; 94US-0188582.

XX 28-JAN-1993; 93US-0013412.

XX 30-JUN-1993; 93US-0087119.

XX (REGC) UNIV CALIFORNIA.

XX Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;

XX Tjian R, Wang E, Weinzierl ROJ;

XX

DR WPI: 1996-333245/33.

XX P-PSDB; AAW06077.

XX Screen for cpds. that bind human TATA-binding protein associated
 PT factor - by testing ability to bind to polypeptide fragments of the
 PT factor, useful as (ant)agonists of transcription factors involved in
 PT disease.

XX Examples; Column 27-36; 86pp; English.

XX This is the nucleotide sequence encoding the Drosophila TATA-binding
 CC protein (TBP) associated factor (TAF) designated TAFII110. The protein
 CC is a component of the TFIIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The encoded protein
 CC has an estimated mol. wt. of 110 kD by SDS-PAGE, a calculated mol. wt.
 CC based on sequence of the gene of 99.4 kD and an estimated pI of 10.1.
 CC The invention relates to purified proteins involved in transcription
 CC by RNA polymerase II, the RNA polymerase which transcribes messenger
 CC RNA. RNA polymerase II transcription proceeds in vitro upon addition
 CC of several nuclear fractions designated TFIIA, B, D, E, F, H, I and J
 CC to RNA polymerase II holoenzyme. Fraction TFIIID has been shown to
 CC contain a TBP and other TAFs. Purification of TFIIID and separation of
 CC its components reveals 7 proteins ranging in size from 30-250 kD.
 CC Serum raised against the TFIIID fraction allowed cloning of the corresp.
 CC genes from lambda-gtl1 expression libraries.

XX Sequence 4615 BP; 1394 A; 1199 C; 1091 G; 931 T; 0 other;

Query Match 5.6%; Score 142.2; DB 17; Length 4615;
 Best Local Similarity 50.7%; Pred. No. 4.1e-29;
 Matches 395; Conservative 0; Mismatches 378; Indels 6; Gaps 2;

QY 1553 GATCAGGATGACATCAATGATGTGACTTCTATGCGAGGGGTCAACCTTAATGAAGAAAT 1712
 DB 2548 GGTGACGATGATATCAACGATGTTGCCGCCATGGGAGGTGTTAACTTGGCGGAGGTCG 2607
 QY 1713 GCCTGCATCTTAGCAACAACTCTGAATGGTTGGCACACTCATCTCACTCATGTAAGAT 1772
 DB 2608 CAGCGAATTTCTCGGCTGTGA---CCGAAACATCGCGACGAGATTCGATCTCTGCAAGAT 2664
 QY 1773 GAACCATTTCTTTTATTGGAGCTCTACAAAAGAGAACTTTAGACATCTGGTAAAGAT 1832
 DB 2665 GAGGTTTTCTTAATCTCCCTCGCTGCACTAGTAATACGGGCAATTAATCTCGAGGCG 2724
 QY 1833 GACATTACAGAACTTAATCTGATGCTGCTGAATCTCCCAAGCAACACAGGAAGCA 1892
 DB 2725 GGACTGGATGAGCGCTCGCAGGATGTGCGCGTCTTGATATCGCAGCGCTGTGAGGCGC 2784
 QY 1893 CTACGAGGCTTCTAGAAAACTGACTGCAATTTGCTCAGCATCGAATCACTACTTACAAG 1952
 DB 2785 CTGAAGAACATCGTTGAGAACTTGGCTGTGATGCGGAGCACCGCATTTGATGTCATCAAG 2844
 QY 1953 GCAAGTGAATAATTAATCTGCTGTAGTATACACAGTCCAGCTCAGCTCAAAATTTCTGAAAG 2012
 DB 2845 TTGGATCCACGCTATGATGCCCGCCCAAGGATGTGCGGCTGAGATCAAGTTTCTCGAGGAG 2904
 QY 2013 CTGATCAATTTGGAGAACGAGAGAAAGATTTGGAAGAAAGAGAAATGTTTAAAGGCA 2072
 DB 2905 CTGGACAAGGCCGAGCAGAGACGACACGAGGAACTGGAGCGTGTGATGCTGCTGCGGCA 2964
 QY 2073 GCCAAGAGTCTTCTTAATAAGAGATCCAGAACAGCTGAGATTAAGCAGAAAGCCAAA 2132
 DB 2965 GCCAAGTCCAGGCTCGAGGTTGGAAGATCCCGAGCAGAGCAAGATGAAGCGGAGGCCAAG 3024
 QY 2133 GAGTTACAGCAATTTGGAACCTTGCACAGATACAGCATAGAGACGCTTAATCTCACAGCTCTT 2192
 DB 3025 GAGATGCAACCGCGCGAATGGAGGAGTTGGCTCAACGAGATGCCAATCTGACGCGCTG 3084
 QY 2193 GCAGCTATTTGGACCAAGAGAGACCACTAGAAATTTGAGGGCTTAAAGAC 2252
 DB 3085 CAGCGATTTGGACCTCGGAAAAAGCTGAAGCTGGAGCGGCAACAGTCAGTTCTGGGAGCG 3144

QY 2253 AACCTTCTTCTCTGG---GACATCCAGCCTGACAGCCACCAACAGTTGATCGTCCA 2309
 Db 3145 GGTTCAGTGGCGGAGTGTGTAAGCAGCTCGGATCTGCGCCGACGAGCTTACGGCCT 3204
 QY 2310 AGAATCAGGAGATCTGCTCAGGGACTTGTATTTTGTATGGAACAGGAGGAGATG 2369
 Db 3205 CGCATAAACGTGTCGAACCTGGCGACATGCTCTTCTACATGAGCAAGCGGGAGTTC 3264
 QY 2370 AAGTATTCGAGCTCTATACCTGGCCCTCTGAAGTACCACCTCCACCTCTTCCATCCA 2428
 Db 3265 TGTCGAGTTCATGCTGTGTAAGACATACCTCAAGTGATGCTGCTGTGTCCTCA 3323

RESULT 11
 AAT79604
 ID AAT79604 standard; cdNA; 4615 BP.
 AC AAT79604;
 XX
 DT 08-OCT-1997 (first entry)
 DE TATA-binding protein associated factor, dTAFII110 cdNA.
 XX
 KW TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP;
 KW initiation; ds.
 XX
 OS Drosophila sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 538..3303
 FT /*tag= a
 XX
 PN US5637686-A.
 XX
 PD 10-JUN-1997.
 XX
 PF 28-JAN-1993; 93US-0013412.
 XX
 PR 28-JAN-1994; 94US-0188582.
 PR 28-JAN-1993; 93US-0013412.
 PR 30-JUN-1993; 93US-0087119.
 PR 09-MAY-1996; 96US-0646715.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Comai L, Dynlact BD, Hoey T, Ruppert S, Tanese N;
 PI Tjian R, Wang E, Weinzierl ROJ;
 XX
 DR WPI; 1997-319113/29.
 DR P-PSDB; AAW25028.
 XX
 PT Nucleic acids encoding human TATA-binding protein associated factor
 PT (TAF) peptide(s) - for production of recombinant peptide(s), used
 PT for modulating transcription of TAFs
 XX
 PS Example 1; Column 35-40; 86pp; English.
 XX
 CC AAT79604 encodes Drosophila TATA-binding protein associated factor
 CC (TAF) polypeptide, dTAFII110 (mol. weight 110kD). TAF peptides derived
 CC from dTAFII130 alpha, dTAFII130 beta, dTAFII40, dTAFII60, dTAFII80,
 CC dTAFII110, dTAFII50, and dTAFII250, their human equivalents and
 CC nucleic acids encoding them, are used to modulate transcription,
 CC including transcription initiation. TAFs are nuclear proteins involved
 CC in RNA polymerase I, II and III transcription. The peptides act by
 CC binding to a different TAF, an activator, or TBP (TATA-binding protein)
 CC or competitively inhibiting association of a TAF domain with another
 CC compound, typically a protein like TBP or another TAF, an activator,
 CC or DNA.
 XX
 SQ Sequence 4615 BP; 1394 A; 1199 C; 1091 G; 931 T; 0 other;
 Query Match 5.6%; Score 142.2; DB 18; Length 4615;

Best Local Similarity 50.7%; Pred. No. 4.1e-29;
 Matches 395; Conservative 0; Mismatches 378; Indels 6; Gaps 2;
 QY 1653 GATGAGGATGACATCAATGATGACTTCTATGCGAGGGGTCAACCTTATGAGAAAT 1712
 Db 2548 GGTGAGATGATATCAACGATGTTCGGCATGGGAGTGTAACTTGGCGGAGGATCG 2607
 QY 1713 GCCTGATCTTATGCAACAACTCTGAATGGTTGGCACACTCAATTCAGTCATCTAAAGAT 1772
 Db 2608 CAGCGAATTCCTGGCTGTA---CCGAACAATCGGCACGAGATTCGATCTCAAGAT 2664
 QY 1773 GAACCATTTCTTTTATTTGAGCTCTACAAAAGAGAACTTTAGACATTTGGTAAAAAGCAT 1832
 Db 2665 GAGGTTTTTCTTAATCTCCCTCGCTCAAGCTAGAAATACGGCAATTAATCTCGGAGCG 2724
 QY 1833 GACATTACAGAACTTAATCTGATGCTGTGAATCTTATCTCCCAAGCAACACAGGAACGA 1892
 Db 2725 GGACTGGATGAGCCGTCGCGGATGTGGCGTCTCTATATCGCACGCCCTGTACGAGCGC 2784
 QY 1893 CTACGAGGCTTCTAGAAAACTGACTGCAATTCCTCAGCATCGAATGACTTACTTACAAG 1952
 Db 2785 CTGAAGAATCATGTTGAGAAGTTGGCTGTGATAGCGGAGCAGCCGATTTGATCTCAAG 2044
 QY 1953 GCAAGTGAATTTACATCTCTGTGTAGTATACAGGTACAGCTCAAAATTTCTTGAAG 2012
 Db 2845 TTGGATCCACGCTATGAGCCCGCCAAAGATGTGCGCGGTGATCAAGTTCTTCGAGGAG 2904
 QY 2013 CTGGATCAATTGGAGAGCAGAGAAAGATTTGGAGAAGAGAAATGTTACTTNAAGCA 2072
 Db 2905 CTGGACAAAGCGGACGAGAGCGACAGAGAACTGGACGCTGAGATGCTGCTCGCGGCA 2964
 QY 2073 GCCAAGATCGTTCTTAATAAAGAACTCCAGAACACTGAGATTAAGCAGAAAGCCAAA 2132
 Db 2965 GCCAAGTCACGTCGAGGTTGGAAGATCCGAGCAGCCCAAGATGAAGCGGAGGCCAAG 3024
 QY 2133 GAGTTACAGCAATTTGGAATTTGCACAGATACAGATAGACGCTTAATCTCAGAGCTTT 2192
 Db 3025 GAGATGCAACCGCCCGAAATGGAGGATTTCCGTCAACAGAGATGCCAATCTGACGGCGCTG 3084
 QY 2193 GCAGCTATTGGACCAAGGAAGAGAGACCACTAGATCTGGAAATTCAGGCTTAAAGAC 2252
 Db 3085 CAGCGGATTTGACCTCGGAAAAGCTGGAAGCTGGAGCGGAAACACAGTCAGTTCGGAGCG 3144
 QY 2253 AACCTTCTTCTCTGG---GACATCCAGCCTGACAGCCACCAACAGTTGCTATCGTCCA 2309
 Db 3145 GGTTCAGTGGCGGAGTGTGTAAGCAGCTCGGGATCTGCGCCGACGAGCTTACGGCCT 3204
 QY 2310 AGAATCAGGAGATCTGCTCAGGGACTTGTATTTTGTATGGAACAGGAGGAGATG 2369
 Db 3205 CGCATAAACGTGTCGAACCTGGCGACATGCTCTTCTACATGAGCAAGCGGGAGTTC 3264
 QY 2370 AAGTATTCGAGCTCTATACCTGGCCCTCTGAAGTACCACCTCCACCTCTTCCATCCA 2428
 Db 3265 TGTCGAGTTCATGCTGTGTAAGACATACCTCAAGTGATGCTGCTGTGTCCTCA 3323

RESULT 12
 ABK32842
 ID ABK32842 standard; DNA; 2307 BP.
 XX
 AC ABK32842;
 XX
 DT 23-APR-2002 (first entry)
 DE DNA encoding human homologue of MPT1 antifungal target.
 XX
 KW antifungal; fungal gene transcription; RPC34; POP3; TPA2; NAB2;
 KW MPT1; MTR2; BOS1; POL30; RSA2; SQP1; MTW1; TFB1; SPC98; BFR1; RNA1;
 KW GCD7; SKI6; NIP1; LCP5; NCE103; ORC2; CNS1; YPD1; TIM10; SRB4;
 KW yeast; fungus; ds; gene.
 XX
 OS Homo sapiens.
 XX

PN WO200202055-A2.
 XX 10-JAN-2002.
 PD
 XX 28-JUN-2001; 2001WO-US20592.
 XX 29-JUN-2000; 2000US-215164P.
 PR 10-AUG-2000; 2000US-224457P.
 XX (ANAD-) ANADYS PHARM INC.
 XX Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;
 PI Mendillo M, Moore D, McCooy M, Sanderson K, Haq T, Zhu S, Long F,
 PI Davidov E, Thompson CM;
 XX WPI; 2002-147962/19.
 DR P-PSDB; AAU82954.
 DR
 XX Screening candidate antifungal compound for interaction with essential
 PT protein, modulation of essential protein activity, binding to essential
 PT protein, by contacting protein with test compound and determining
 PT effects
 XX
 PS Disclosure; Figure 80; 522pp; English.
 XX
 CC The invention describes a method of screening a candidate antifungal
 CC compound for interaction with essential proteins (EP) or for modulation
 CC of EP activity e.g fungal gene transcription. The proteins tested in the
 CC invention include RPC34, POP3, TFA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,
 CC SOT1, MTW1, TFB1, SPC98, BFR3, RNAL, GCD7, SKI6, NIP1, LCP5, NCE103,
 CC EGO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans
 CC and human homologues. The method involves contacting a culture with one
 CC or more test compounds and determining the effects on the growth or
 CC viability of the culture of cells which preferably comprises fungal cells
 CC or yeast cells. Preferably the identified compounds interact with, or
 CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor
 CC compounds identified by the method are useful for preventing or
 CC inhibiting fungal, particularly C. albicans growth in culture or in a
 CC mammal. The antifungal agents interact with essential fungal elements
 CC that can be used to treat fungal infection by preventing the growth and
 CC preferentially killing the fungi, but does not inhibit the biological
 CC activity of mammalian homologues. This sequence encodes a target protein
 CC used to test the antifungal compounds, described in the method of the
 CC invention.
 XX
 SQ Sequence 2307 BP; 359 A; 969 C; 724 G; 255 T; 0 other;
 Query Match 5.3%; Score 135.8; DB 24; Length 2307;
 Best Local Similarity 66.7%; Pred. No. 1.8e-27;
 Matches 194; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
 QY 556 TTCTCCGACAAATGCTAGAAATGGAAGAAATGCAAGAACTTCCTTGCATGTTAATAAA 615
 DB 1776 TGGCAGGAAACTATGGAACACGTGGAAGAAATGTAATAATTCCTATCTAGTTAATAAA 1835
 QY 616 ACTAGCATGTAGTGGATCAGACAGTCCCTGAAATGGGCAAAATGTGAGAGCTGGTGA 675
 DB 1836 ACTGGCTTCATCTGGCAGCAGCTCTACAGACAGACAGCTAATGTGAAGAGCTCGTGCA 1895
 QY 676 ACAACTTTTGGATCAAAATTCGAAGCAGAGAAATTTACTAGGAACTGTATGTTGAAC 735
 DB 1896 GAATTTACTGATGGAATAATGAGCAGAGAGATTTTCAACAGCAGGTATATACCGAGACT 1955
 QY 736 CAAGCTTTTCACTCAGGCTCACCCTGGTTCCTTTTCAAGAAACCGTGGTTCCTTACG 795
 DB 1956 TAATTTCTTCACTCAGCTTACCTTGTGCTTTCCTTGAAGAGAGCTTACCCGCTTGAG 2015
 QY 796 ACAACTTCTGCCTTAACCTCCAGAGCTTCATCCCAACATGTGTTCCAGCAGAC 846
 DB 2016 ACAGCTGACCCCGACACCTCCGCGCTTTCATCCAGCAGAGCAGCAGCAGC 2066

RESULT 13

ABN33770
 ID ABN33770 standard; DNA; 60 BP.
 XX
 AC ABN33770;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human spliced transcript detection oligonucleotide SEQ ID NO:6518.
 XX
 KW Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcripome; oligonucleotide library; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200210449-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 20-JUL-2001; 2001WO-IB01903.
 XX
 PR 28-JUL-2000; 2000US-221607P.
 PR 02-MAY-2001; 2001US-287724P.
 XX
 PA (COMP-) COMPUGEN INC.
 XX
 PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 XX
 DR WPI; 2002-257383/30.
 XX
 PT New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes
 XX
 PS Example 1; SEQ ID 6518; 47pp; English.
 XX
 CC The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridizing selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterising the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 60 BP; 17 A; 17 C; 11 G; 15 T; 0 other;
 Query Match 2.3%; Score 60; DB 24; Length 60;
 Best Local Similarity 100.0%; Pred. No. 6.6e-07;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 873 GTACTACACAGTAACTTCTCTCTGTGGTGACAACTACAGTGTCTCAAGCCAGCTG 932
 DB 1 GTACTACACAGTAACTTCTCTCTGTGGTGACAACTACAGTGTCTCTCAAGCCAGCTG 60

RESULT 14


```

Db 3066 GGACTGGATGAGCGCTCGCAGGATGTGGCCGTTCTGTATATCGCAGCCCTGTTCAGGAGGC 3007
Qy 1893 CTACGAGCCCTTCTAGAAAACTGACTGCAATTGCTCAGCATCGAATGACTACTTACAAG 1952
Db 3006 CTGAAGAACATCGTTGAGAAGTTGGCTGTGTAGCGGAGCACCGCATTGATGTCATCAAG 2947
Qy 1953 GCAAGTGAATAATT 1965
Db 2946 GTTTGTTGGAAT 2934

```

Search completed: February 16, 2003, 22:28:41
 Job time : 631 secs

This Page Blank (uspto)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 20:54:50 ; Search time 16.4197 Seconds
(without alignments)
2152.161 Million cell updates/sec

Title: US-09-763-909-2
Perfect score: 4264
Sequence: 1 GTLVTKVAPVSAPPKVVSSCP.....KZQHERALFTIRTLTLTY 852

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|--------------|---------------------|
| 1 | 3990 | 93.6 | 801 | 1 T2DT_HUMAN | Q92750 homo sapien |
| 2 | 1307.5 | 30.7 | 1083 | 1 T2D3_HUMAN | Q00268 homo sapien |
| 3 | 775 | 18.2 | 921 | 1 T2D3_DROME | P47825 drosophila |
| 4 | 239.5 | 5.6 | 2035 | 1 HFCL_HUMAN | P51610 homo sapien |
| 5 | 238 | 5.6 | 2090 | 1 HFCL_MESAU | P51611 mesocricetu |
| 6 | 220 | 5.2 | 1367 | 1 AMYH_YEAST | P08640 saccharomyc |
| 7 | 219 | 5.1 | 5179 | 1 MUC2_HUMAN | Q02817 homo sapien |
| 8 | 213.5 | 5.0 | 5376 | 1 ZAN_MOUSE | O88799 mus musculu |
| 9 | 206.5 | 4.8 | 865 | 1 CPN_DROME | Q02910 drosophila |
| 10 | 205.5 | 4.8 | 1161 | 1 DAN4_YEAST | P47179 saccharomyc |
| 11 | 203.5 | 4.8 | 1322 | 1 YAG3_YEAST | P39712 saccharomyc |
| 12 | 201 | 4.7 | 725 | 1 AGA1_YEAST | P32323 saccharomyc |
| 13 | 199 | 4.7 | 2700 | 1 ZAN_HUMAN | Q9Y493 homo sapien |
| 14 | 197.5 | 4.6 | 881 | 1 YJH8_YEAST | P47033 saccharomyc |
| 15 | 197.5 | 4.6 | 1609 | 1 FIG2_YEAST | P25653 saccharomyc |
| 16 | 193.5 | 4.5 | 670 | 1 VG50_HSV1 | Q00130 ictaluriid h |
| 17 | 191.5 | 4.5 | 1260 | 1 ALS1_CANAL | P46590 candida alb |
| 18 | 191 | 4.5 | 797 | 1 VGLX_HSV1 | P28598 equine herp |
| 19 | 187.5 | 4.4 | 3726 | 1 TRX_DROME | Q02659 drosophila |
| 20 | 186 | 4.4 | 3178 | 1 YS89_CAEEL | Q09624 caenorhabdi |
| 21 | 184.5 | 4.3 | 1537 | 1 FLO1_YEAST | P32768 saccharomyc |
| 22 | 182 | 4.3 | 1075 | 1 FLO5_YEAST | P38894 saccharomyc |
| 23 | 182 | 4.3 | 3866 | 1 HRX_MOUSE | P55200 mus musculu |
| 24 | 181.5 | 4.3 | 662 | 1 MUC1_XENLA | Q05049 xenopus lae |
| 25 | 181.5 | 4.3 | 1199 | 1 N121_RAT | P52591 rattus norv |
| 26 | 181 | 4.2 | 1858 | 1 P3K2_DICDI | P44674 dictyostelli |
| 27 | 177.5 | 4.2 | 1140 | 1 YM96_YEAST | Q04893 saccharomyc |
| 28 | 176 | 4.1 | 2541 | 1 TALI_HUMAN | Q9Y490 homo sapien |
| 29 | 175.5 | 4.1 | 2090 | 1 N214_HUMAN | P35658 homo sapien |
| 30 | 175.5 | 4.1 | 3969 | 1 HRX_HUMAN | Q03164 homo sapien |
| 31 | 173.5 | 4.1 | 1802 | 1 HKR1_YEAST | P41809 saccharomyc |
| 32 | 173 | 4.1 | 2404 | 1 SON_MOUSE | Q9qk47 mus musculu |
| 33 | 172 | 4.0 | 886 | 1 VGP3_EBVA8 | Q07284 epstein-bar |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 172 | 4.0 | 1018 | 1 HMW1_MYCPN | O50365 mycoplasma |
| 35 | 172 | 4.0 | 1119 | 1 ALS3_CANAL | O74623 candida alb |
| 36 | 172 | 4.0 | 1189 | 1 YK82_YEAST | P36170 saccharomyc |
| 37 | 171.5 | 4.0 | 1780 | 1 YK26_CAEEL | P34333 caenorhabdi |
| 38 | 170 | 4.0 | 1306 | 1 MSB2_YEAST | P32334 saccharomyc |
| 39 | 169.5 | 4.0 | 388 | 1 MPT1_YEAST | P50105 saccharomyc |
| 40 | 169.5 | 4.0 | 745 | 1 OCT1_PIG | O29076 sus scrofa |
| 41 | 169.5 | 4.0 | 2426 | 1 SON_HUMAN | P18583 homo sapien |
| 42 | 169 | 4.0 | 1229 | 1 N121_HUMAN | Q9y2h3 homo sapien |
| 43 | 166.5 | 3.9 | 743 | 1 OCT1_HUMAN | P14859 homo sapien |
| 44 | 165 | 3.9 | 606 | 1 SP2_HUMAN | Q02086 homo sapien |
| 45 | 165 | 3.9 | 2541 | 1 TALI_MOUSE | P26039 mus musculu |

ALIGNMENTS

RESULT 1

T2DT_HUMAN
ID T2DT_HUMAN STANDARD; PRT; 801 AA.
AC Q92750;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor TFIIID 105 kDa subunit (TAFII-105)
DE (TAFII105) (fragment).
GN TAF4B OR TAF2C2 OR TAFII105.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE:97011146; PubMed:8858156;
RA Dikstein R., Zhou S., Tjian R.;
RT "Human TAFII 105 is a cell type-specific TFIIID subunit related to hTAFII130".
RL Cell 87:137-146(1996).
CC -1- FUNCTION: CELL TYPE-SPECIFIC SUBUNIT OF TFIIID THAT MAY FUNCTION AS A GENE-SELECTIVE COACTIVATOR IN CERTAIN CELLS. TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS.
CC -1- SUBUNIT: TFIIID IS COMPOSED OF YATA BINDING PROTEIN (TBP) AND A NUMBER OF TBP-ASSOCIATED FACTORS (TAFs). TAFII105 MAY FORM HETERODIMERS WITH TAFII130.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FOUND IN B CELLS BUT NOT IN OTHER TESTED CELLS.
CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Y09321; CAA70499.1; -
Genew; HGNC:11538; TAF4B.
MIM: 601689;
InterPro: IPR003894; TAF_hom.
SMART; SM00549; TAFH; 1.
Transcription regulation; Nuclear protein.
FT NON_TER 1
SQ SEQUENCE 801 AA; 85658 MW; D12B4932FEA9CD2 CRC64;

Query Match 93.6%; Score 3990; DB 1; Length 801;
Best Local Similarity 99.9%; Pred. No. 5.1e-199;
Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTLVTKVAPVSAPPKVVSSGPELPAPVAVKAPNTTIIQPPANQLPLPGTVLIKSNSGPL 60
|||||

Db 1 GTLVTKVAPVSAPPKVSQSPRLPAPQIVAVKAPNTTTTIOFPANLQLPGCTVLKNSGSL 60
 QY 61 MLVSPQOTVTRAEFTSNITSRPAVPANPOTVKTCTVPNSSSOLIKKAVTPVKKLAQIGT 120
 Db 61 MLVSPQOTVTRAEFTSNITSRPAVPANPOTVKTCTVPNSSSOLIKKAVTPVKKLAQIGT 120
 QY 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTTLKPSLIGASSTPSNEPNLKAENSA 180
 Db 121 TVVTVTPKSSVQSVAVPTSVTVTPGKPLNTVTTLKPSLIGASSTPSNEPNLKAENSA 180
 QY 181 VQINLSPMLENVKCKNFELAMLIKACSGSPENGVNKKVLQGLLDKAEAEFTTK 240
 Db 181 VQINLSPMLENVKCKNFELAMLIKACSGSPENGVNKKVLQGLLDKAEAEFTTK 240
 QY 241 LVEKSSQPHLVPFLPKSVVALRQLPNSQSFQCCVQOQSSDMVIACTCTTTTSPV 300
 Db 241 LVEKSSQPHLVPFLPKSVVALRQLPNSQSFQCCVQOQSSDMVIACTCTTTTSPV 300
 QY 301 VTTTSSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTLHSGVPTAATGGTTA 360
 Db 301 VTTTSSSSSEKSIIVSGATAPRTVSQTLNPLAGPVGAKAGVTLHSGVPTAATGGTTA 360
 QY 361 GRLGLQTSKPLVTSVANTVTVTSLOPEKPVSGTAVTSLPVTGCTSGAAICLPSVKP 420
 Db 361 GRLGLQTSKPLVTSVANTVTVTSLOPEKPVSGTAVTSLPVTGCTSGAAICLPSVKP 420
 QY 421 VVSFCDWHICKPVGIPVQIKLAQCPVLSPAGIPTSSSSKQLFSLFHVQOQSGNPK 480
 Db 421 VVSFCDWHICKPVGIPVQIKLAQCPVLSPAGIPTSSSSKQLFSLFHVQOQSGNPK 480
 QY 481 QVTTTSHSTLTATOKGOKTQVNTIIPTSQPPASILKQITLPGNKKILSLQASPTQNR 540
 Db 481 QVTTTSHSTLTATOKGOKTQVNTIIPTSQPPASILKQITLPGNKKILSLQASPTQNR 540
 QY 541 IKENVTSRDEDDINDVTSMAGVNLNEENACILATNSLVTGLQSCDEFFLFGALQ 600
 Db 541 IKENVTSRDEDDINDVTSMAGVNLNEENACILATNSLVTGLQSCDEFFLFGALQ 600
 QY 601 KRILQIGKHDITENSADVNLISQATQRLRGLLEKLTATIAQHRMTYKASENVILCSD 660
 Db 601 KRILQIGKHDITENSADVNLISQATQRLRGLLEKLTATIAQHRMTYKASENVILCSD 660
 QY 721 QHRDANLTAAAGPRKKRPLESGIEGLKNDLLASGTSLSLTATKQHRPRITRICLDLI 780
 Db 721 QHRDANLTAAAGPRKKRPLESGIEGLKNDLLASGTSLSLTATKQHRPRITRICLDLI 780
 QY 781 FCMEQEREMKYSRALYLALK 801
 Db 781 FCMEQEREMKYSRALYLALK 801

RESULT 2

T2D3_HUMAN
 ID T2D3_HUMAN STANDARD; PRT: 1083 AA.
 AC 000368; Q99721; Q9BX42; Q9BR40;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)
 DE (TAFII135) (TAFII-130) (TAFII130).
 GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97336072; PubMed=9129867;
 RA Mengus G., May M., Carre L., Chambon P., Davidson I.;

Human TAF(II)135 potentiates transcriptional activation by the AP-2s of the retinoic acid, vitamin D3, and thyroid hormone receptors in mammalian cells.;
 Genes Dev. 11:1381-1395(1997).
 [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beare D.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Levaslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConnachie L.J., Mclay K., McMurray A.A., Milne S.A., Misty D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhan R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 105-1083 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=97098442; PubMed=8942982;
 RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
 RT "Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100.";
 RT Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
 CC -1- FUNCTION: MAKES PART OF TFIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS AND REPRESSORS. POTENTIATES TRANSCRIPTIONAL ACTIVATION BY THE AP-2S OF THE RETINOIC ACID, VITAMIN D3 AND THYROID HORMONE.
 CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE TAF2C FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; Y11354; CAA72189.1; -
 DR EMBL; AL137077; CAC36006.1; -
 DR EMBL; AL109911; CAC23212.2; -
 DR EMBL; U75308; AAC50901.1; -
 DR TRANSFAC; T02328; -
 DR Genew; HGNC:11537; TAF4.
 DR MIM; 601796; -
 DR InterPro; IPR003894; TAF_hom.
 DR SMART; SM00549; TAFH; 1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 39 42 POLY-HIS.
 FT DOMAIN 52 57 POLY-ALA.
 FT DOMAIN 98 101 POLY-GLY.
 FT DOMAIN 142 148 POLY-ALA.
 FT DOMAIN 268 275 POLY-PRO.

FT DOMAIN 331 337 POLY-ALA.
 FT DOMAIN 680 683 POLY-PRO.
 FT DOMAIN 808 813 POLY-ALA.
 FT DOMAIN 828 831 POLY-ASP.
 FT CONFLICT 105 117 PGPPSPRRPLVPA -> GRGLLQOGRGREG
 (IN REF. 3).
 FT CONFLICT 136 136 A -> S (IN REF. 2).
 FT CONFLICT 185 185 G -> GPG (IN REF. 2).
 FT CONFLICT 233 264 MISSING (IN REF. 3).
 FT CONFLICT 293 P -> L (IN REF. 3).
 SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

Query Match 30.78; Score 1307.5; DB 1; Length 1083;
 Best Local Similarity 40.08; Pred. No. 2.8e-200;
 Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;

QY 13 PKRVSSG-----PRLPAPQIVAVKAPNTTTTQFPANLQLPPTGVLIKNSGSLM 61
 DB 414 PTATTSIGIRATLTPTVLAPRLPQP-----PQNFTNIQ---NFQLPPGMVLRSNGQLL 464
 QY 62 LVSPQOQVTR-----AETTSNITSRAVPANPOTVTKICTVPSNSSLIKKVAVTPVKKL 115
 DB 465 MI-PQALQAQQAHAQAPOTMAPRPAATPTSAPPVQISTVQAPGTPTIAR-QVTP----- 518
 QY 116 AQIGTIVVTVPKPSVQSVAVPTSVVTVTPGKPLNTWT--TLKPPSSLGASS-----TPS 168
 DB 519 ----TTIKQV---SQATTVQPSATQPSRQVQVQVLVGLGAATASLGATATVQTGTGP 571
 QY 169 NE-PNLKAENSAVQIINLSPTMLNENKCKNFKLAKLKGSGSQSPQEMGONVKLVQEQ 227
 DB 572 RTVPGATTTSSAAVE-----TMENVKCKNFKLTLKLGSSGKQSTETAAVVKELVQNL 625
 QY 228 LDKAIEAEFTKRLVYELKSPQPHLVFLLKSVVALRQLLPNSQSFQICQVQ-----TS 283
 DB 626 LDGKIEAEFTSRLYRELNSPQVLYVFLKRLSPALRQLTPDSSAFTQQSQOQPPPPPTS 685
 QY 284 SDMVIATCTTIVTSPVVTITVSSQSEKSTIIVSGATAPRTVSQVTLNPLAGPVGKAGV 343
 DB 586 Q-----ATTALFVNLSSSVQRTAGTATVTSALOPPVLSL----- 722
 QY 344 VTLHSVGPTAATGTTAGTGLLOTSKPLVTSVANTVTVSLQPKPVVSGTAVTLSLPAV 403
 DB 723 -----TQTPQGVGQGGQPTPLVIQ-----QPKP-----GALIRPQV 756
 QY 404 TFGTSGAAICLPVKPVVSCWDHCKPVGTGTVQIKLQAPGVLVQAPIGTSGSSKQ 463
 DB 757 TLAQT-----PWVALRQPH-NRIMLTTPQIQI----- 783
 QY 464 LFLSLFHVVPQSGGNEKQVTTISHSSTLTQKCGQKMPVNTIIPPTSPASILKQITL 523
 DB 784 -----NPLQPVVVKPAVL 797
 QY 524 PGNKIL---SLQASPTQKNRIKENVTSFCRDEDDINDVTSNAGVNLNEENACILATNSEL 580
 DB 798 PGTALSAVSAQAAAKNKLKPGGSGFRDDDDINDVASMAGVNLSEARILATNSEL 857
 QY 581 VGTLIQSKDEPFIFGLAQRIIDIGKHHDTIELNSDANVLISQATQERLGLLEKITA 640
 DB 858 VGTITRSCKDETFLLQAPLQRRILEIGKHGITEHPDWVSYSVSHATQORLQNLVEKISE 917
 QY 641 IAOHRMTYTKASENYILCSDTRSQLKLEKLDQEKQKDLEREMLLKAAKSRNKEDP 700
 DB 918 TAOQKNFSYKDDRDYEQASDVRAQLKFEQDQIEKQKDKQREILMRAAKSRNQEDP 977
 QY 701 EQLRKQAKELQOLELAQIOHRDANLTATAAIGPRKKRPLE-----SGIEGLKDNLLAS 755
 DB 978 EQLRLKQAKEMQOQELAQRDRDANLTALAAIGPRKKRVKDYCPGSGABSGPGSVVP 1037
 QY 756 GTSSLTATKQLHRPRTICRLDILFQWQREMKYSALYLALK 801
 DB 1038 GSSGVGTPTQRTQRTIRVRLDLIFCLENERETSHSLLYKAPFLK 1083
 "The genome sequence of Drosophila melanogaster."

RESULT 3
 T2D3_DROME
 ID T2D3_DROME STANDARD; PRT; 921 AA.
 AC P47825; P49845; Q9VUY7;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription initiation factor TFIID 110 kDa subunit (P110)
 DE (TAF11-110) (110 kDa TBP-associated factor).
 GN TAF110 OR CG5444.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 519-540; 597-616 AND 857-874.
 RC TISSUE=Embryo;
 RX MEDLINE=931145326; PubMed=7678780;
 RA Hoey T., Weinzierl R.O.J., Gill G., Chen J.-L., Dynlacht B.D.,
 RA Tjian R.;
 RT "Molecular cloning and functional analysis of Drosophila TAF110
 RT reveal properties expected of coactivators.";
 RL Cell 72:247-260(1993).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 398-406; 520-540 AND 860-877.
 RC TISSUE=Embryo;
 RX MEDLINE=93317591; PubMed=8327460;
 RA Kokubo T., Gong D.-W., Roeder R.G., Horikoshi M., Nakatani Y.;
 RT "The Drosophila 110-kDa transcription factor TFIID subunit directly
 RT interacts with the N-terminal region of the 230-kDa subunit.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5896-5900(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.G., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Arif J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Klumpp B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou S., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A
 CC CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS
 CC AND REPRESSORS. MAY FUNCTION AS A COACTIVATOR BY SERVING AS A SITE
 CC OF PROTEIN-PROTEIN CONTACT BETWEEN ACTIVATORS LIKE SPI (OR BTD)
 CC AND TFIIID COMPLEX.
 CC -!- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L06861; -; NOT_ANNOTATED_CDS.
 DR EMBL; S63550; AAB27433.1; -;
 DR EMBL; AB003528; AAF49536.1; -;
 DR TRANSFAC; T02121; -;
 DR FlyBase; FBgn0010280; Tafi110.
 DR InterPro; IPR003894; TAF_hom.
 DR SMART; SM00549; TAFH; 1.
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 66 82 POLY-GLN.
 FT DOMAIN 108 111 POLY-GLN.
 FT DOMAIN 259 265 POLY-GLN.
 SQ SEQUENCE 921 AA; 93338 MW; 27E685289872767 CRC64;
 Query Match 18.2%; Score 775; DB 1; Length 921;
 Best Local Similarity 29.6%; Pred. No. 7.2e-33;
 Matches 258; Conservative 128; Mismatches 308; Indels 178; Gaps 26;
 QY 25 PQIVAVKAPNTTIQPPANLQPPGTVLTKNSGPMML-----VSPQQTIVTRAETS 76
 DB 133 POSPSTLSTLNTGQTPA-----LLVKTNGFOLLRVGTTGPTVTQTITNTSNN 184
 QY 77 NITSRAVPANPQTVKICVPSNSSQ-----LIKKVAVTPVKKLAIGTIVTTVTP 127
 DB 185 NTSITNHTTTQ-IRLQVPAASMTNTATSNIVNSVASSGVANSQPHLTQLNAQ 243
 QY 128 KPSSVQSVAVPTSVVTVTPGKPLNTVTLKPSLSGASPTSPNEPNLKAENSAVQINLSP 187
 DB 244 APQLPOITQITIPAOOSOOQOVNYSAGGTATAVSSTA-----ATT 287
 QY 188 TMLNVK-KCKNFLAMLIKAGSQSPENGVKVLVEQLLDKAEAEFTKLYVELK 246
 DB 288 TQGNTEKCKRFLANLIEL--STREPKPVEKNVRLTQELVNAVPEPEFCDLRLLEN 345
 QY 247 SSPQHLVPLKSVVALROL-----LPNSQSFTQ--- 276
 DB 345 ASPQCLIGFLKSLPLLRQALYTKELVIEGPKPPQHVGLAGLSQQLPKQAIRPG 405
 QY 277 ----OCVQTSDDWJATCTTTTTPVTVTTVSSQSEKSIIVSGATAPRTVS---VQT 329
 DB 406 PSQTTTIGTQVRMI--TPNALGTPTPTTIGHTTISKQPN---IRLPTAPRLVNTGGIRT 460
 QY 330 LNPAGPVCAKAGVTLHVSFGTATGTTAGTGLLOTSKPLVTSVANTVTVTSLOPEKP 389
 DB 461 QIP-SLQVPGQANVOIR--GPQHAQLQRTGSVQIRATRRP-----PNSVPTAN----- 506
 QY 390 VVSGTAVTSLPAVTFGETSGAICLPSVKPVVSCWDHICKPVIGTPVQIKLAOPGVL 449
 DB 507 -----KLAVKVGQTIKAI-TPSLHP-----PSLAISGGP-----PPTPL 543
 QY 450 SQPAGITGSSSK---QLFSLFVVVQPSGGNEKQVTTISHS----- 489
 DB 544 SVLSTLNSASTTTLPIPSLPTVHLPEALRAREQMSLNHNHFDKALVEIKAPSLHP 603
 QY 490 -----TLTIQCGKQTPVNTIITSPPPASILKQITLPGNKILSLQ-----AS 534

RESULT 4

HFCL_HUMAN
 ID HFCL_HUMAN STANDARD; PRT; 2035 AA.
 AC P51610;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFCL) (VCAF)
 DE (CCF).
 GN HFCL OR HCF1
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 CC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND PARTIAL SEQUENCE.
 RX MEDLINE=93272419; PubMed=8392914;
 RA Wilson A.C., Lamarco K., Peterson M.G., Herr W.;
 RT "The VP16 accessory protein HCF is a family of polypeptides processed
 RL from a large precursor protein.";
 RL Cell 74:115-125(1993).
 RN [2]
 RP SEQUENCE OF 65-2035 FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Fetal brain;
 RX MEDLINE=95130107; PubMed=7829097;
 RA Frattini A., Faranda S., Redolfi E., Zucchi I., Villa A.,
 RA Patrosso M.C., Strina D., Susani L., Vezzoni P.;
 RT "Genomic organization of the human VP16 accessory protein, a
 RL housekeeping gene (HCF1) mapping to Xq28.";
 RL Genomics 23:30-35(1994).
 RN [3]
 RP PROCESSING.
 RX MEDLINE=96033796; PubMed=7590226;
 RA Wilson A.C., Peterson M.G., Herr W.;
 RT "The HCF repeat is an unusual proteolytic cleavage signal.";
 RL Genes Dev. 9:2445-2458(1995).
 CC -!- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
 CC TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
 CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
 CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
 CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.
 CC -!- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
 CC 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N-
 CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
 CC NONCOVALENTLY, ASSOCIATED.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING. IN THE
 CC SHORTER FORM THE N- AND THE C-TERMINAL FRAGMENTS FAIL TO

CC ASSOCIATE.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL TISSUES AND THE
CC ADULT KIDNEY. PRESENT IN ALL TISSUES TESTED.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL.
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PPCE/THET, WITHIN THE HCF REPEAT.
CC -1- PTM: GLYCOSYLATED; CONTAINS O-LINKED N-ACETYLGLUCOSAMINE RESIDUES.
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L20010; ; NOT_ANNOTATED_CDS.
CC EMBL: X79198; CAA55790.1; AUT_INIT.
CC Genew: HGNC:4839; HCFCL.
CC MIM: 300019; ;
CC InterPro: IPR003961; FN.III.
CC InterPro: IPR001798; Kelch.
CC Pfam: PF00041; fn3; 1.
CC Pfam: PF01344; Kelch; 5.
CC SMART: SM00060; FN3; 1.
CC Nuclear protein; Repeat; Alternative splicing; Glycoprotein.
CC
CC FT REPEAT 44 89
CC FT REPEAT 93 140
CC FT REPEAT 148 194
CC FT REPEAT 217 265
CC FT REPEAT 266 313
CC FT DOMAIN 1010 1439
CC FT REPEAT 1010 1035
CC FT REPEAT 1072 1097
CC FT REPEAT 1101 1126
CC FT REPEAT 1158 1183
CC FT REPEAT 1286 1311
CC FT REPEAT 1314 1339
CC FT REPEAT 1349 1374
CC FT REPEAT 1414 1439
CC FT VARSPLIC 382 450
CC FT CONFLICT 564 584
CC FT CONFLICT 603 603
CC FT CONFLICT 1164 1164
CC FT CONFLICT 1873 1873
CC SEQUENCE 2035 AA; 208841 MW; 99207FBE875204C0 CRC64;

Query Match 5.6%; Score 239.5; DB 1; Length 2035;
Best Local Similarity 22.6%; Pred. No. 9.4e-05;
Matches 164; Conservative 89; Mismatches 245; Indels 229; Gaps 38;

QY 2 TLTKVAPVSAPPKVSRRPLP-----APQIVAKPNTT-----IQ 39
Db 446 TLTPQAAP--APTPTTIQVLPVPOSSISVPPAARTQGVPAVLKVTGPQATTGTPLVTM 503

QY 40 FPNAN-----LQLPPGVLII---KNSGGLMLVSPQOT-----VTRAETTSNI--TS 80
Db 504 RPASQAGKAPVTYSLPAGVVMVPTQSAQGTIGSSPQMSGMAALAAAAATQKTPSS 563

QY 81 RP---AVPAN-----PQTVKICTVP-----NSSOLIKKVVAVTPVKKLAQ 117
Db 564 RPTVLSVPAGTTIVKTMVTPGTTLTPATVKVASSPVMVSNPATRMLKATA-----AQ 616

QY 118 IGTTV-----VTVPKPSVQSVAVPTSVTVTPKPLNTVTLK--PSSL--GASS 165
Db 617 VGVSVSSATNTSRPTIIVHKSGTV--TVAQQAQVTVTVVGVGVTKITLVKSPISVPGGSA 675

QY 166 TPNENPL-----KAENSAAVQINLSPTMLENVKCKNPL--AMLIKACSGSQSP--- 214
Db 676 LISNLKVMVSVTKPQVTSVAVTQASTGPVTQIQTKGLPAGTLKLVTSADGKPTTI 735

QY 215 -----EMGVNKKVLVEQLLDAKIEAEFFTKLYVELKSSPQPHLVPLPKSVVALRQLLP 269
Db 736 ITTQASGAGTKPTILGI-----SSVSPSTT---KPGTTTIKTIP 773

QY 270 NSOSFTQCCVQQTSSDMVIAFCTTTVT-----SPVVT-----TTVSSSQSEKSI 314
Db 774 MSALITQAGATGVTSPGKISPIITIKVMTSGTGAPAKIITAVPKIATGHGQGGVTVQ 833

QY 315 IVSGA-----TAPRT-----VSVQTLNPLAGPVGAK--AGVVTLHRSVGPAAATG 356
Db 834 VLKAGAPQPCITLRTVPMGCVRLVTPVTVSAVKPAVTVLVKGTGTGTLGTGTGTVST- 892

QY 357 GTTAGTGLLTSLPLVTSVA--NFTVTVSLQPEKPVVSGTAVTSLSPAVTGETSGAIC 414
Db 893 -SLAGAGGHTSASLATPITTLGTIATLSQ----VINPTAITVSAQTTLTAAGGLTTP 947

QY 415 LPVVKPVVSCWDHICKPVIGTVPQIKL-AQPGVLSQPA-GIPTGSSSKQLSLFHVVQ 472
Db 948 TITWQPV-----SQPTQVTLITAPSGVEAQPVDLPVS-----ILA 983

QY 473 QPSGGNEKQVTTISHSTLTIQKCGQKTMPTVNTIIPTSQPPASI-----LKQITLPG 525
Db 984 SP-----TTEQPTATVIADSGQGVQPGVTVLVCNPPCETHETGTNTATTIVA 1035

QY 526 NKILSLQASPTQRNRIKENVTSFRDEDDINDVTSMAGVNLNEENACILATNSLVGTLI 585
Db 1036 N--LGGHPQPTQVQFV-----CDRQEAASLVSTVG---QON-----GSV 1072

QY 586 QSCKDEP 592
Db 1073 RVCNPP 1079

RESULT 5
HCFCL_MESAU STANDARD; PRT; 2090 AA.
ID HCFCL_MESAU
AC F51611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Host cell factor c1 (HCF) (VP16 accessory protein) (HFC1) (VCAF) (CFF)
GN HCFCL.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabepu Y., Nishimoto T.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 kDa AND A MINOR 300 kDa POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PPCE/THET, WITHIN THE HCF REPEAT (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D45419; BAA08258.1; -
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 5.
DR SMART; SM00060; FN3; 2.
KW Nuclear protein; Repeat.
FT REPEAT 44 89
FT REPEAT 93 140
FT REPEAT 148 194
FT REPEAT 217 265
FT REPEAT 266 313
FT DOMAIN 1010 1448
FT REPEAT 1010 1035
FT REPEAT 1072 1097
FT REPEAT 1101 1126
FT REPEAT 1157 1182
FT REPEAT 1295 1320
FT REPEAT 1323 1348
FT REPEAT 1358 1383
FT REPEAT 1423 1448
SQ SEQUENCE 2090 AA; 214942 MW; E495E8B1F2385E17 CRC64;

Query Match 5.6%; Score 238; DB 1; Length 2090;
Best Local Similarity 21.7%; Pred. NO. 0.00012;
Matches 147; Conservative 84; Mismatches 219; Indels 226; Gaps 32;

QY 10 VSAPKPVSSGRLPAQIVAVKAPNTTIIQFPANLQPLPGTVLIKNSGPMVLSPQQT 69
DB 537 IGSNFMQSMGMAAALAAATQKIPSSA-----PTVLSPAGTIVT-----VAVTPGTT 588
QY 70 TRAEITSNITSPVPANPQT-----VKICTVPNSSQ-----LIKKVAVTPVKRL 115
DB 589 LPA--TVKVASSPMVSNPATRLMTAAQVCTSVSSAANTSTRILIVHKSGETVVAQ 646
QY 116 AQIGTVV-----TVVPKP-----SSVOSVAVPTS----- 140
DB 647 AQVTVTVGVGKTKITLVKSPISVPGSALISNLGKVMVSVQTKVQTSVATGQSTGPV 706
QY 141 -----VVTVTPGRPLNTVTL-----KPSLSGAST-----PSNE 170
DB 707 TQIIQKGLPLAGTILKLVTSADGRTTIIITQAGSAGSRPTIIGISVSPSTTKPGTT 766
QY 171 PNLKAENSAVQINLSPITMLNKKCKNFKLAML-IKLACGSSQSPMGQNVKKLVEQLLD 229
DB 767 TIIKIPMSAIIQAGATGVTSTPGIKSPITITTKVMTSGTAP-----AKIITAV-- 818
QY 230 AKIEAEETRKL-YVELKSSP-QPHLVFELKKSVALRQLLPNSOSFIOQCQVQSTSDMV 287
DB 819 PKIATGHGQGGVTVLKGAPGQG-----AILRTVP-----MSGVR 855
QY 288 IATCTTTTTSVPTVTVSSQSEKSIIVSGATAPRTVSQTLNPLAGPVGAKGVVTLH 347
DB 856 LVTPVTVSAVKPVT-----LVVKGTT-----GVTLG 884
QY 348 SVGPTAATGGTGTAGLLQTSKPLVTSVA--NTVTVLSLQPEKPVVSGTAVTSLPAVTF 405
DB 885 TVTGTVST--SLAGAGHASTASLATPITTLGTIATLSQ-----VINPITVSAQTTL 938
QY 406 GETSAAATCLPSVKPVVSCWDHICKVIGTPVQKL-AQGPVLVSQA-GIPTGSSSKQ 463
DB 939 TAAGGLTPTITMQPV-----SQPTQVTLITAPSGVEAQPVHDLPSV----- 980
QY 464 LPSFLHVVOQPGGNEKQVTTISHSTLTIOKCGOKTTPVNTIIPTSQFPASII----- 517
DB 981 -----ILASP-----TTEQPTATVTIADSGGQVQPGTIVLVCSNPPCETHETGTT 1026

QY 518 -LKQITLPQNKILSLQASPTQKNRIKENVTSCFRDEDDINDVTSMAGVNLNENACILAT 576
DB 1027 NTATTIVAN--LGGHPQPTQVQFV-----CDRQAASLVTSVAG-----QQN----- 1068
QY 577 NSELVGTLIQSKDEP 592
DB 1069 -----GNVVRVCSNPP 1079

RESULT 6
ID AMYL_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucanase SI/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN STAI OR STAZ OR MAL5 OR YI019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Chentles C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAI".
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STAZ and SGA genes
RT from Saccharomyces cerevisiae".
RL FEBS Lett. 239:179-184(1988).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -|- SIMILARITY: TO S.POMBE SPBC215.13.
CC -|- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38061; CAA86176.1; -
DR EMBL; M16164; AAA35014.1; -
DR EMBL; M16165; AAA35015.1; -
DR EMBL; X13857; CAA32069.1; -
DR PIR; B26877; B26877.
DR PIR; A26877; A26877.
DR PIR; S48478; S48478.
DR SGD; S0001458; MUC1.
KW Hydrolase; Glucosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.

| | | | | |
|----|----------|----------|-----------------------------------|---------------------|
| FT | DOMAIN | 3893 | 4928 | VNFD 17 (PARTIAL). |
| FT | DOMAIN | 4029 | 4148 | VNFD 18 (PARTIAL). |
| FT | DOMAIN | 4149 | 4263 | VNFD 19 (PARTIAL). |
| FT | DOMAIN | 4264 | 4283 | VNFD 20 (PARTIAL). |
| FT | DOMAIN | 4384 | 4503 | VNFD 21 (PARTIAL). |
| FT | DOMAIN | 4504 | 4623 | VNFD 22 (PARTIAL). |
| FT | DOMAIN | 4624 | 4743 | VNFD 23 (PARTIAL). |
| FT | DOMAIN | 4744 | 4863 | VNFD 24 (PARTIAL). |
| FT | DOMAIN | 4864 | 5261 | VNFD 25. |
| FT | DOMAIN | 5259 | 5295 | EGF-LIKE. |
| FT | DISULFID | 5263 | 5274 | BY SIMILARITY. |
| FT | DISULFID | 5268 | 5283 | BY SIMILARITY. |
| FT | DISULFID | 5285 | 5294 | BY SIMILARITY. |
| FT | CARBOHYD | 339 | 339 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 499 | 499 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 1216 | 1216 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 1239 | 1239 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 1314 | 1314 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 1814 | 1814 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 1908 | 1908 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 1933 | 1933 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 2028 | 2028 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 2111 | 2111 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 2142 | 2142 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 2332 | 2332 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 2533 | 2533 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 2575 | 2575 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 2692 | 2692 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 2812 | 2812 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 3052 | 3052 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 3065 | 3065 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 3144 | 3144 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 3172 | 3172 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 3288 | 3288 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 3292 | 3292 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 3782 | 3782 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 4005 | 4005 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 4136 | 4136 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 4243 | 4243 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 4254 | 4254 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 4335 | 4335 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 4376 | 4376 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 4586 | 4586 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 5136 | 5136 | N-LINKED (GLCNAC..) |
| FT | CARBOHYD | 5252 | 5252 | N-LINKED (GLCNAC..) |
| SQ | SEQUENCE | 5376 AA; | 579908 MW; 0844DB7DF2A2620 CRC64; | (POTENTIAL).. |

Query Match 5.0%; Score 213.5; DB 1; Length 5376;
 Best Local Similarity 23.0%; Pred. No. 0.0075;
 Matches 157; Conservative 89; Mismatches 260; Indels 177; Gaps

| | | |
|----|--|----------------------|
| Qy | 8 APVSAP--PKVSSGRLPARQIVAVKAPTNTTTQFPA---- | NQLPPGTVLIKNSGPLM 61 |
| Db | 551 SPLLPPTGPSESTVPLTLMSEQTSPTKATTVTIETPTTPEEATIPETTTPVTE--- | VI 607 |
| Qy | 62 LVSPQQVTTRAETSNTISRPVPANPQTVKICTVPSNSSQLKKVAVTPVKKLQIGIT 121 | |
| Db | 608 NVSPKETSIPPEVT-----IPTVEIIV-----SPEELISPTVEPTVPDVTAAVY 652 | |
| Qy | 122 VWTIV-BKPSSV-QSAVPTSVMVTTPGKPLNTVTLTKPSSL-GASSTPSNEPNLKAEN 177 | |
| Db | 653 EATNASPEETSVPPEVILTVEVTVSPDE--TVTPTEVPVILEATAFPGETTLYTEVP 710 | |
| Qy | 178 SAAVOINLSPTMLBNVKKCNFLAMLIKACSQSOSPQMCKVKKLVLEQLLDAA---IEA 234 | |
| Db | 711 TVPTEVTVGVHTEVTV-----SPE-----ETSVPTEESTISTEVTVSP 748 | |
| Qy | 235 EEFRTRK-----LVVELKSSPQPHL-----VPFLKKSVALRQLLPNSQSIQCVOOTS- 283 | |
| Db | 749 EETTVPTTEVPVILEATASTPGTELTLYTEVPTVPTTEVTVGVHTEVTVNSP-----BETSV 802 | |
| Qy | 284 -SDMVATCTTIV-----TTSPVVWTVTVSSSQSEKSIIVSGATAPRTVSOTLNLAGPVG 338 | |

```

DB 803 PTEETISTEVTVSPDETTLPTEVTVSTEVTVNS-----PETSVPPEETI----- 849
QY 339 AKAGVTVLHVSQGTAAAGTTAGTGILL-----QTSKPLVTSVANTVTVTS-----LQPEK 388
DB 850 -----LTLTVTEVTVPTVTEVTVHTEVTVNSPEETSVPTEETISTEVTVSPDETTLPTEV 905
QY 389 PVVSGTAVTSLPAPVTFGTSGAAICLSVKP-----VVSFQWDHICKVIGTPVQIK 441
DB 906 PTVS-TEVTVNSPEET-----SVPEETILTITVSPDETVFPTEGTTL----- 949
QY 442 LAQPGVLSQAPAGI---PTGSSSKOLFSLFHVQVQPSGSGNEKQVTTISHSS----- 490
DB 950 ---PTEVTVTVTEVTVTFGTET---VPTVTVTVSTEMGVTEVTVTVPEETSITVETAT 1004
QY 491 -----LTIQKCGQKTMVNV-TIPTSQFPP---ASILKQIPLPKNKILS 530
DB 1005 VLPASIPPEETTTPTVETTTTPEETTPAEVTVTVPPASIPPEETASLTVETTTTPEETT 1064
QY 531 ---LQASPTQKNRIKENVT-----SCFDEDDI-NDVTSMAG--VNLNEENACILIATNSE 579
DB 1065 PTEVTVTVPEKTIPTVETTVTVPPASIPPEETTVPEETTIASEETTVSTQETTLTQESA 1124
QY 580 LVGTILI---QSCKDEPFLFAGL 599
DB 1125 VTQTSIACRPPCPSPPLMPIGPL 1147

RESULT 9
CPN_DROME STANDARD; PRT; 865 AA.
ID CPN_DROME Q02910;
AC 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE Calphotin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
OF CA+2 PER MOL OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
COMPOUND EYES AND OCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
DEVELOPMENT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

CC -----L02111; AAA28405.1;
DR EMBL; L02111; AAA28405.1;
DR PIR; A47282; A47282
DR FlyBase; FBgn0010218; Cpn.
KW Calcium-binding.
FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFLICT 43 43 I -> T (IN REF. 2).
FT CONFLICT 64 64 I -> V (IN REF. 2).
FT CONFLICT 76 76 T -> A (IN REF. 2).
FT CONFLICT 100 100 P -> PP (IN REF. 2).
FT CONFLICT 126 127 VO -> AP (IN REF. 2).
FT CONFLICT 154 154 I -> V (IN REF. 2).
FT CONFLICT 160 160 S -> T (IN REF. 2).
FT CONFLICT 534 534 A -> E (IN REF. 2).
FT CONFLICT 699 699 I -> T (IN REF. 2).
FT CONFLICT 703 703 V -> L (IN REF. 2).
FT CONFLICT 721 721 D -> E (IN REF. 2).
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;

Query Match 4.8%; Score 206.5; DB 1; Length 865;
Best Local Similarity 21.7%; Pred No. 0.0015;
Matches 122; Conservative 82; Mismatches 242; Indels 115; Gaps 22;

QY 1 GTLVTKV-APVSAP-----PKVSSGRLPAP-----QIVAVKAPNTTTIQFPANLQ 45
DB 4 GTIPSPVAPVAAPVTPSAVAPVQVSPAAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 63
QY 46 LPPTVILKSNGLMLYSQOTVTRAETTSNITSRPAVPANPQVTKICTVPNS-SSQIL 104
DB 64 IPAPAPIAAASVTPVASVAP--PVVAAPT-----PAASPVSTPVAQIDPVASAPVA 115
QY 105 KKVAVTPVKKL-AQIGTTVTVTVPKPSVQSVAVPTSVVT-VTPGKPLNTVTLKPSLSG 162
DB 116 PPVAATPTPVQIPVAAPVIAATPPVAASAPTAAVTPVISPVIAFPVVPVPAATTPVVAAP 175
QY 163 ASSTPSNEPNLKAENSAVQINLSPTMLE-----NVKKCNFLAMLIKACS 209
DB 176 VAAVPAAPVPAVPLAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 227
QY 210 GSQSPKMGONKVLVEQLLDKIAEAEETKRLVYELKSSQPHLPFLKKSVALRQLLP 269
DB 228 -----PEVSVVATKPLAAAEPPVVVAPPATPTPVVAPAAASPHVSVAPAVETAVVA 278
QY 270 NSQSFIOQCQQTSSDMVIACTTTTTPSVVTVTVSSQSEKSIIVSGATAPRTVSVOT 329
DB 279 VSAS-----TEPPVAAATLTITAPET-PALAPVVAESQVAANTVATPPTP-APEPET 328
QY 330 LNPLAGPVGAKAGVTVLHVSQGTAAAGTTAGTGILLQTSKPLVTSVANTVTVTSLOPEK 389
DB 329 IAP-----PVAETPEVASVAVAEETPPVVPVPA-----AESIPAP 364
QY 390 VVSGTAVTSLPAPVTFGTSGAAICLPSVKPVVS-----FCWDHICKPVICT 436
DB 365 VVATTPVPATL-AVTDPDVDTASAV--PELPPVIAAPSPVPSAVAETPVDLAPVLPVPAAE 421
QY 437 PVQIKLQAGPVLSQAPAGIPTGSSSKOLFSLFHVQVQPSGSGNEKQVTTISHSSTLTIOK 496
DB 422 PVPVVAEETETPAPASAPVITIALDPEVAPVIAAASDAPAE-----APSAAPAPI---- 473
QY 497 GQKTMVNTIPTSQFPPASI 517
DB 474 -VSTPTTASVPETTAPPAV 493

RESULT 10
DAN4_YEAST STANDARD; PRT; 1161 AA.
ID DAN4_YEAST
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.

```


| | | | | | |
|---|-------------|---|---------------------------------------|------------------------|--------------|
| FT | CARBOHYD | 464 | 464 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 509 | 509 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 554 | 554 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 599 | 599 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 644 | 644 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 689 | 689 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 734 | 734 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| FT | CARBOHYD | 888 | 888 | N-LINKED (GLCNAC. . .) | (POTENTIAL). |
| SQ | SEQUENCE | 1322 | AA; 138072 | MW; AAFD1FF13267CEA | CRC64; |
| Query Match | | | | | |
| Best Local Similarity 4.8%; Score 203.5; DB 1; Length 1322; | | | | | |
| Matches 142; Conservative 114; Mismatches 276; Indels 215; Gaps 31; | | | | | |
| QY | 22 | LPAPQ-IVAKAPNTTITQF---- | PANLQLPEGTVLKNSG----- | PLMLVSPQOT | 68 |
| DB | 663 | LPTDETIIVIRPTTATTAMITQPNNDTFTSTSTETITVTGNTGLPTDETIIVIRPTT | 722 | | |
| QY | 69 | VTRAETT----- | SNITSRAVPANPQNTVKICTVPNSSOLIKKVAIVPVK | 113 | |
| DB | 723 | ATTAMTTTQPNNDTFTSTETITVTGNTGVPD-ETVIVIRPTSE----- | 768 | | |
| QY | 114 | KLAQIGTTVTVTPKPSVQSVAVPTSVVTPPGKPLN-TVTLK-PSLGLASSTPSNEP | 171 | | |
| DB | 769 | -----GLISTTPTWTGTTSTETITVTGNTGQPTDETIVIRPTSEGLVTT-TTEP | 822 | | |
| QY | 172 | -----NLKAENSAVQINLSPITMLENVKCKNFAMLIK----- | LACSGSQSPQMGNVKK | 222 | |
| DB | 823 | WTGTFSTSTETITGNTGQPT----- | DETVIIVKPTTATLSSSSSQ----- | 869 | |
| QY | 223 | LVEQLLDKAKIEBEETRKLYELKSSPOHLPVFKKSVVALRQLLPNSQSFIOCVQOT | 282 | | |
| DB | 870 | -----ITSFITSARPIITPPYPS----- | NGTSSVSSSVISS | 900 | |
| QY | 283 | S--SDMVIATCTT--VTSPVTTTVSSSQSEKIIIVSGATAPTVSVQTLNPLAGPVG | 338 | | |
| DB | 901 | SDTSLVSSSVTSLVSSPVISSFTSP----- | VISSTTSALSSES----- | 946 | |
| QY | 339 | AKAGVTVLHSGVPTAATGTTAGTGLQTSKPLVTSVA--NTVTTVSLQPEKPVVSGTAV | 396 | | |
| DB | 947 | SKSSVPTSS--STGSGSESETGSASSASSSSSSISSESPKSTYSSSLPPVTSATTSQEI | 1004 | | |
| QY | 397 | TLSLPAVTFEGTSGAICLPVKPVVFCWDHICKPVGTGV----- | QKLAQPGVLSQP | 452 | |
| DB | 1005 | TSSLVPVTTTSEQTTLV-----TVTSCSHVCTESISSAIVSTATVTVSGATTEYTTW | 1059 | | |
| QY | 453 | AGIPGTSSSKOLFSLFHVYVQOPSGNGEKQVTTISHSPTLTQKC----- | GQ | 498 | |
| DB | 1060 | CPISTETIKQ---TTETTKOTKGTTEOTTETTKOTTVTTISSCESDVCSKTASPAIVST | 1116 | | |
| QY | 499 | KTMPVNTI-----IPTSOFPASILKQITLPGNKILSLQASPTQKNRIKENVTSCF | 549 | | |
| DB | 1117 | STATINGVTTEVTTWCPISTTESKQOTLVTVTSCGVCSETTSPA---IVSTATA-- | 1170 | | |
| QY | 550 | RDEEDINDVTSMAGVNLNEENACILATNSLVGLIQCKDEPFLFIALQ----- | 600 | | |
| DB | 1171 | -----TVNDVTVTVSTWRPQ----- | TTNEQSVSSKMNSTATSETTTNTGAAETTTSTGAEE | 1220 | |
| QY | 601 | ----- | KRLIDGKKHHDITELNSDAVNLSIQATQERL | 631 | |
| DB | 1221 | TKTVTSSISRENHAETASATDVIGHISSVSVSVSETGNTKSLTSSGLSTMSQ--QPRS | 1278 | | |
| QY | 632 | RGLEKL-TATAQHRMTTYKASENYIL | 657 | | |
| DB | 1279 | TPASSMVGSSASLEISTYAGSANSLL | 1305 | | |
| RESULT 12 | | | | | |
| ID | AGAL_YEAST | STANDARD; | PRT; | 725 | AA. |
| AC | P32323; | | | | |
| DT | 01-OCT-1993 | (Rel. 27, Created) | | | |
| DT | 01-OCT-1993 | (Rel. 27, Last sequence update) | | | |

| | | | | | |
|--|--|-----------------------------------|-----|---|----------------------|
| DT | 01-OCT-1996 | (Rel. 34, Last annotation update) | | | |
| DE | A-agglutinin attachment subunit precursor. | | | | |
| GN | AGAL OR YNR044W OR N3431. | | | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomyces. | | | | |
| OX | NCBI_TaxID=4932; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=91304412; PubMed=2072914; | | | | |
| RA | Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.; | | | | |
| RT | "The AGAL product is involved in cell surface attachment of the | | | | |
| RT | Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin."; | | | | |
| RL | Mol. Cell. Biol. 11:4196-4206(1991). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | Pohl T.M.; | | | | |
| RL | Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases. | | | | |
| CC | -!- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ | | | | |
| CC | SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY | | | | |
| CC | CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ, | | | | |
| CC | RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR | | | | |
| CC | AGGREGATION DURING MATING. | | | | |
| CC | -!- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO | | | | |
| CC | A CORE SUBUNIT. | | | | |
| CC | -!- AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE). | | | | |
| CC | -!- PTM: EXTENSIVELY O-GLYCOSYLATED. | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |
| CC | ----- | | | | |
| DR | EMBL; M60590; AAA34382.1; - | | | | |
| DR | EMBL; 271659; CAA96325.1; - | | | | |
| DR | PIR; S17031; S17031. | | | | |
| DR | PIR; A41258; A41258. | | | | |
| DR | SGD; S0003327; AGAL. | | | | |
| KW | Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat; | | | | |
| FT | Phenome response. | | | | |
| FT | SIGNAL | 1 | 22 | POTENTIAL. | |
| FT | CHAIN | 23 | 725 | A-AGGLUTININ ATTACHMENT SUBUNIT. | |
| FT | DOMAIN | 53 | 493 | 2 X APPROXIMATE REPEATS. | |
| FT | REPEAT | 53 | 149 | 1-1. | |
| FT | REPEAT | 395 | 493 | 1-2. | |
| FT | DOMAIN | 182 | 307 | 18 X APPROXIMATE TANDEM REPEATS, SER/THR- | |
| FT | REPEAT | 182 | 188 | RICH. | |
| FT | REPEAT | 189 | 195 | 2-1. | |
| FT | REPEAT | 196 | 202 | 2-2. | |
| FT | REPEAT | 203 | 209 | 2-3. | |
| FT | REPEAT | 210 | 216 | 2-4. | |
| FT | REPEAT | 217 | 223 | 2-5. | |
| FT | REPEAT | 224 | 230 | 2-6. | |
| FT | REPEAT | 231 | 237 | 2-7. | |
| FT | REPEAT | 238 | 244 | 2-8. | |
| FT | REPEAT | 245 | 251 | 2-9. | |
| FT | REPEAT | 252 | 258 | 2-10. | |
| FT | REPEAT | 259 | 265 | 2-11. | |
| FT | REPEAT | 266 | 272 | 2-12. | |
| FT | REPEAT | 273 | 279 | 2-13. | |
| FT | REPEAT | 280 | 286 | 2-14. | |
| FT | REPEAT | 287 | 293 | 2-15. | |
| FT | REPEAT | 294 | 300 | 2-16. | |
| FT | REPEAT | 301 | 307 | 2-17. | |
| SQ | SEQUENCE | 725 | AA; | 73353 | MW; 70420C853B0B01F8 |
| | | | | | CRC64; |
| Query Match 4.7%; Score 201; DB 1; Length 725; | | | | | |
| Best Local Similarity 20.4%; Pred. No. 0.0023; | | | | | |

Matches 144; Conservative 120; Mismatches 309; Indels 134; Gaps 30;

```
QY 4 VTKVAPVSAPKVKSSGRPLPAQIVAVKAPN-----TTTIQPPANLQ-----LPPGTGLI 53
Db 18 LTNIALASDPETI-----LVTITKNDANGVVVTTVS-PALVSTSTIVQAGTTTL 66
QY 54 KNSGGPLML-----VSPO-----QTVRAET--TSNITSRPVAPNP--QTVKICTVP 97
Db 67 YTWCPPLTVTSAAELSPISVATLUSRFSTLTLTEVCHACFPSSSLPPTTLTUSVTS 126
QY 98 NSSSLIKKVVAVPVKLAQIGTVV--TTPKPSVQSVAVPTSVV-TVTPGKPLMTVT 154
Db 127 KFTSYICPTCHTTAISLSLSEGVTTVVSSAIEPSSAIIISPVTLSLSTSNP--TTT 184
QY 155 TLAPSSLGASSTPSNPNLKAENSAVQINLSTMLNFKCNFLAMLKLCAGSSQSP 214
Db 185 SLSTSTSPSST-STPSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSP 228
QY 215 EMGQNVKKLVQLLDKAEAEFRKLYVELKSSPQPHLVFPLKKSVALRQLLPNSOF 274
Db 229 SSTSTSSSLTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSPSKST 282
QY 275 IOOCVOOTS-----SDMVIATCTTVTTPVTTTSSSQSEKSIIVSGATAPRTVSVQT 329
Db 283 SASSTSTSYSTSTSPSLTSSPTLASTSP-SSTSISSSTFTDSTSLGSSIASSTSVSL 341
QY 330 LNPAGPVGAKAGVVLHSGVPTAAGCTTACGLTQTSPLVT--SVANTVTVTSLOPE 387
Db 342 YSP-STPVSVPS--TSSNVATPSMTSSTVETTVSSQSSSEYITKSSISSTIPFSFMSY 398
QY 388 KPVVSG--TAVTSLPVTGTEGSAICLPSPKPVVFCWDHICKPVICTPVQIKLAQP 445
Db 399 FTVVSGVTTWYTWCPYSSSESTSTLTMHETVTTDMATVCTHESCMP----- 445
QY 446 GPVLSPAGIPTGSSSKQLFLFHVQOPSGGNEKQVTTISHSTLTIQCGOKTTPVNT 505
Db 446 ----SQTSLTSS-----IKMSTKNVATSVSTSTVSSVACSTCAETSHSYSS 490
QY 506 IIPSTOPPPASILKQITLPCNKILSQASPTQKNRI---KENTVSPRD--EDDINDVTS 560
Db 491 VQPASS---SSVTQOTSTKSWSSMTSDENFKHATGKYHYVTSSTSTSVSSEATS 547
QY 561 MAGVN--LNEENACILAT---NSELVGLTIQSCDPEFLFALOKRIIDICKKHDI 614
Db 548 TSSIDESQSQSHSLTSTSVLSSSLSATL---SSDSTILLFSSVSLSYE---QSPVTT 601
QY 615 LNSDAVNLSIQATQOERLGLLEKLTATAQHRMTYKASENYILCSDT 661
Db 602 LQISSTSEILQPTSS-----TATATISASTSLSATSTSTPST 639
```

RESULT 13

```
ZAN_HUMAN STANDARD; PRT; 2700 AA.
ID ZAN_HUMAN
AC Q9Y493; O00218;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin (Fragment).
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE-99018118; PubMed-9799793;
RA Glockner G., Scherer S., Schattevoy R., Boright A., Weber J.,
RA Tsui L.C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes."
RL Genome Res. 8:1060-1073(1998).
```

```
[2]
RN SEQUENCE OF 2338-2700 FROM N.A.
RP TISSUE=Testis;
RC MEDLINE-97271566; PubMed-9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RL zonadhesin gene (ZAN)."
CC Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHUS.
CC -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -!- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC -!- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```

```
FT CARBOHYD 2505 2505 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 2374 2375 NNQKWA -> RAGPGP (IN REF. 1).
FT NON_TER 2700 2700
SQ SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;

Query Match
Best Local Similarity 4.7%; Score 199; DB 1; Length 2700;
Matches 147; Conservativity 91; Mismatches 268; Indels 160; Gaps 34;

QY 1 GTLVTKV--APVSAPP-----KVSSGRLPAPQIV-AVKAPNTTII 38
DB 372 GTMLELLGSPAGSPGLPKRVGSRPYQWNTSVTPSGHQPLFKGQGSNTASV 431
QY 39 QPANQLPPGTVLKSNGPLMLVSPQOTVTRAETTSNITSIRPVPANPQVACTV 98
DB 432 VAMGFILINPGTQPKVKV-LPELPPVSPVSSGTGSETT-GLTENPTISTKKPT 483
QY 99 SSSOLIKKVAIVPVKLAQIGTVTTVPKPSVQSVAVPTSVVIVTPGKPLNTVTL 158
DB 484 -----IEKPSVTEK-----PTVPK-----EKPTIPEKPTISTEKP--TIPSEKP 522
QY 159 SSLGASSTPSNPNLKAENSAVAQINLSPTMLENVKKCKNFAMLIKACSGSQSP 218
DB 523 -----NMSEKPTIPSE-----KPTILTPEKTPIS-----EKPTIPSEKPTI-- 559
QY 219 NVKKLVQLLDKABEEPTKRLKVELKSSPOPHLVFLLKKSVALRQLLPNSQSF 276
DB 560 STEKPTVTEPTTPEETT--TYME-----EP-VIPEKPSIPEKPSIPEKPTISME 611
QY 277 QCVQOTSSDMVIATCTTTTTSPPVTTTVSSSQSEKSIIVSG-----ATAPRTVS-----V 327
DB 612 EYIISTEKPTICEKPTIPTPEKPTISTEKSIPKPTTPEKPTIPEKPTISTEKP 671
QY 328 QTLNLAGP--YKAGAGVVTLHSVGTGAATGTTAGTGLLOTSKPLVTSVANTVT--VS 383
DB 672 PTEKPTISPEKLTIPTEKLTIPTEKPT-----ISTEPTTTEETIISTEPRS 726
QY 384 LQEPKPVSGTAVTLSLPAVTGETSGAAICLPSPVKPVVSCFWDHICRPVIGT-----P 437
DB 727 IPMEKPTLTPEETTTTSVEETII--STEKLTIPMEKPTIS-----TEKPTIPTPEKPTISP 778
QY 438 VOI-----KLAQPG-----PV-----LSQAPAGITGSSSKQLSIFHVHQQPSGN 478
DB 779 EKLTIPTEKLTIPTEKPIPIEETIISTEKLIPTEKPTISTEKTIS-----TEKPTIPT 834
QY 479 EK-----QVTTTSHSTLTIOCKGQKTMVNTIIP-----SQFPPASILKQITLPGNKI 528
DB 835 EKPTIPTETTTIS-TEKLTIPTEKPTISPEKLTIPTEKPTISTEKTIPTEKLTIPTEK 892
QY 529 LSLQASPTQKNRIKENVTSCFDEDDINDVTSMAGVNLEE-----NACILATNSLVG 582
DB 893 ----PTIPTPEKPTIPTKLTALRPHPSPTATGTLAALVMSHAPSPMTSVILGTTTTSRS 949
QY 583 TLIOQSC 588
DB 950 STGMSC 955

RESULT 14
YJH8 YEAST
ID YJH8 YEAST STANDARD; PRT; 881 AA.
AC P47033;
DT 01-FEB-1996 (Rel. 33, Created)
DR 01-FEB-1996 (Rel. 33, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 89.2 kDa protein in SCPl60-SMC3 intergenic region.
GN YJL078C OR J1027
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-S288c;
RX MEDLINE=96093911; PubMed=7483841;
RA Miosga T., Schaff-Gerstenschlaeger I., Chalwatzis N., Baur A.,
RA Boles E., Fournier C., Schmitt S., Velten C., Wilhelm N.,
RA Zimmermann F.K.;
RT "Sequence analysis of a 33.1 kb fragment from the left arm of
RT Saccharomyces cerevisiae chromosome X, including putative proteins
RT with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain
RT and a putative alpha 2-SCB-alpha 2 binding site.";
RL Yeast 11:681-689(1995).
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z49353; CAA89370.1; -
CC EMBL: X83502; CAA58492.1; -
CC EMBL: X88851; CAA61314.1; -
CC HSSP: P04284; ICPE.
CC
CC SGD: S0003614; YJL078C.
CC InterPro: IPR001283; Allrgn_V5/Tpx1.
CC PRINTS: PR00837; V5TPXLIKE.
CC ProDom: PD000542; Allrgn_V5/Tpx1; 1.
CC SMART: SM00198; SCP; 1.
CC PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
CC PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
CC KW Hypothetical protein.
CC FT DOMAIN 166 379 ALA/SER/THR-RICH.
CC SQ SEQUENCE 881 AA; 89152 MW; 37DBAC660CA9D12A CRC64;

Query Match
Best Local Similarity 4.6%; Score 197.5; DB 1; Length 881;
Matches 145; Conservativity 87; Mismatches 268; Indels 189; Gaps 27;

QY 47 PGTGTVL-----IKNSGPLMLVSPQOTVTRAETTSNITS--RPVAPNAPQV 91
DB 146 PPGNLTGFEAEVEPELIISTSSSS-----SSSSTSTSDTSTVSTISSIMPV-AQGYTT 199
QY 92 KICTVPNSSOLIKKVAIVPVK--KLAQIGTVTVVTPKP-----SSVQSVAVPTSVVT 143
DB 200 TVSSAASSSS--LKSTTINPAKTALTASSSVITTSSTESVGSSTVSSASSSVTTSYAT 257
QY 144 ---VTPGKPLNVTTLKPSLGLASSTPSNEPNLK-----AENSAVQINLSPTMLE 191
DB 258 SSSTVVSDDATSTTTTSSVATSSSTSSDPTSSATAASSSDPSSAAAS---SSASTE 314
QY 192 NVYKCKNFIA-----MLIKACSGSQSPGEMQNVK-----KLVEQ 226
DB 315 NAASSSSAISSSSSMVSAPLSLTLTSTASSRSVTSNSVNSVFANTTVFSAQTTSSVA 374
QY 227 LLDKATAEFEETKLVVELKSSPQPH--LVFPLKKSVVALROLLPNSQS-----FIOQCQ 280
DB 375 SLSSVAADDIOGTSKATSVSEHTSIVTSATNAQYATRLGSSSSSSGAVSSSVA 434
QY 281 QTSSDMVIA-----TCITTVTTSPPVT--TTVSSSQSEKSIIVSQA 319
DB 435 QSVLSNVIAVNTDVSVTSVSSSTAHTTKDTATTSVITASESITSETAQOASSSTEKNISNA 494
QY 320 TAPRTVSVQTLNPLAGPVGAKAGVTVLHSVGTATGTTAGTGLLOTSKP----- 370
DB 495 T-----SSSIYSNASVSGHVTYAAEYAITSEQSALATSPATNCSSIVK 541
QY 371 -----LVTSVANTVTTVSLQPEKPVVSGTAVTSLPAV----- 403
DB 542 TTLENSSTTTTATATKSTTTLATTANNSTRAATAVTID--PTLDPDTSASPTDNKHTS 600
```

```
QY 404 TFGETSGAAICLPSPKVPVSEFCWDHICKPVIGTPVQIKLAOPGVL----- 449
Db 601 TYGSSGCA-SLSLRKTTTISVSNTQIVSTCTSDSDSPFALSPATTESNLIT 659
QY 450 -----SQAPIGPTGSSKOLFSLFHVVOQPSGGNEKQVTTISHSSTLTITQKCGQKTMP 502
Db 660 NTITASCSTDSNFTSAASDTEAFRTISTS-----CSTLNGASTQT-SELTTSMPK 712
QY 503 VNTLIPTSOFPASILKQITLPGNKILSLQASPTQKNRI-----KENVTSCFRDEDDI 555
Db 713 TMTVVPASSPFTT-----TTCLENDNDTAFSSIXTEVNAATLIINPGTSSLASDFATSEKP 768
QY 556 NDVTSMAGVNLNEENACILATNSELVGLT 584
Db 769 NEPTSVKSTS-NEGTSSTTTTYQQTATL 796

RESULT 15
FIG2 YEAST
ID FIG2.YEAST STANDARD; PRT; 1609 AA.
AC P25653;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Factor induced gene 2.
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92397594; PubMed=1523889;
RA Wilson C.; Grisanti P.; Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RL cerevisiae chromosome III contains two new open reading frames.";
RL Yeast 8:569-575(1992).
CC -1- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC -1- INDUCTION: BY MATING PHEROMONES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59720; CAA42254.1; -
DR PIR; S19504; S19504.
DR PIR; S25345; S25345.
DR SGD; S0000685; FIG2.
SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;

Query Match 4.6%; Score 197.5; DB 1; Length 1609;
Best Local Similarity 21.1%; Pred. No. 0.01;
Matches 166; Conservative 106; Mismatches 327; Indels 187; Gaps 34;

QY 2 TLVTKVAPVSAP-PKVSSGPRLPAPQIVAVKAPNTTTIQPPANLQPLPPGVTLIKNSGPL 60
Db 799 TVITSVTATSTIPLLSTSSVLSSTVSEGAKNPAASEVINTQV-SATSEATSTSTQV 857
QY 61 MLVSPQOVTVAETTSNITSPAVPANDQIVKICTV-----PNSSQLIKKVAIVPVKKL 115
Db 858 SATSATATASESTTSQVSTASETISTLGTQNTFTTGTLLFPALSTEMINTTVYS--RKT 915
QY 116 AQIGTIVTVTPKPSVQSVAVPSTVVTGPKPLNTVTLKPSLSGASPPSPNEPLKA 175
Db 916 LIISTEVCS-----HSKCVPVITEVVTSK-----GTPSNHGSQT 951
QY 176 ENSAAVQINLS--PTMLENVKCKKNFL---AMLIKACSGSQSPM-----GQNVKKLVE 225
Db 952 LQTEAVEVTLSSHQVTMTSTEVCSNICTPTVITVSQMRSTPPFPYLITSSSTSSSLASTKK 1011
```

```
QY 226 QLLDAKIEAEFEETKRLKVELKSSPOPHLYPEL---KKSVVALROLLPNSQSFIOQCVOQT 282
Db 1012 SSLEASSEMSTFS---VSTQSLP---LAFTCEKRSTTSVSQW---SNTVLNTNTIMSS 1060
QY 283 SSDMVIATCTTIVTITSP-----VTTTVSSSQ-----SEK 312
Db 1061 SSVVISINERKPSSTSPYNFSSGYSLPSSSTPSPSYSLSTATTINGIKTVTTWCPLAEK 1120
QY 313 SIIVSGATAPRTV-----SVO-TLNPLAGVPVGAAGVVTUHLHSGVPTAAT 355
Db 1121 STVAASSQSSRSVDRFVSSSKPSSLSQTSIQYTLSTATTISGLKTVTTWCPLTSKST 1180
QY 356 GGTAGTGLLOTSKPLVTSVAN-TVTVTSLOP-----EKPVVSGTAVTLSLPAV 403
Db 1181 LGATTQTS--STAKVITTSASSATSTISLSTSTSESSSGYLSKGVCSTGECTQDVPTQ 1238
QY 404 TFGETSGAAICLPSPKVPVSEFCWD-----HICKPVIGTPVQIKLAOPG--PVLSQ 451
Db 1239 SSSPASTLAYS-PSVSTSSSSSFSTTTASTLTSTHTSVPLLPSSSSISASSPSTSLST 1297
QY 452 PAGIPTGSSSKOLFSLFHVVOQPSGGNEKQVTTISHSSTLTITQKCGQKTMVNTIIPTSQ 511
Db 1298 SLSPAFATSS-----TLPTATAVSSSTFIASSLPLS-SKSSLSLSPVSSSILMSQ 1346
QY 512 FPPASILKQ--ITLPGNKILSLQASPTQKN-RIKENVTS-----CFRDEDDI---- 555
Db 1347 FSSSSSSSSSLASLP-----SLSISPTVDTVSVLQPTTSIATLTCTDSQCOQEVSTICNG 1401
QY 556 ---NDVTSMAGVNLNEENACILATNSELVGLTIOCKDEPFLFICALQKRLDIG----- 607
Db 1402 SNCDVTSATTPPSTVTDVMTCTGSECQKTTSSSC-DGYSCKVSEYTKSSATISACSGE 1460
QY 608 --KKHDITELNSDAVNLISQATQERLRL-----LEKLTAAQHRMTTYKASENYILCS 659
Db 1461 GCQASATSELNSQVVTMTSVITPSAITTSVEVHSTESTISITTVKPVYTS-----S 1513
QY 660 DTRSQL 665
Db 1514 DTNGEL 1519
```

Search completed: February 16, 2003, 21:55:46

Job time : 42.4197 secs

This Page Blank (uspto)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2003, 21:59:29 ; Search time 21.8929 Seconds
(without alignments)
1145.042 Million cell updates/sec

Title: US-09-763-909-2

Perfect score: 4264

Sequence: 1 GTLVTKVAPVSAPPKVSSGP.....KZHQHERALFTIRTLTLTY 852

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents.AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|-------|--------------------|
| 1 | 3990 | 93.6 | 801 | 1 | US-08-725-012-2 |
| 2 | 1307.5 | 30.7 | 737 | 1 | US-08-188-582-16 |
| 3 | 1307.5 | 30.7 | 737 | 1 | US-08-646-715-16 |
| 4 | 775 | 18.2 | 921 | 1 | US-08-188-582-2 |
| 5 | 775 | 18.2 | 921 | 1 | US-08-646-715-2 |
| 6 | 237 | 5.6 | 2035 | 1 | US-08-046-585-5 |
| 7 | 237 | 5.6 | 2035 | 1 | US-08-393-703-5 |
| 8 | 237 | 5.6 | 2035 | 5 | PCT-US93-11721-5 |
| 9 | 198 | 4.6 | 862 | 1 | US-08-325-267A-4 |
| 10 | 197 | 4.6 | 2972 | 4 | US-09-579-181-2 |
| 11 | 197 | 4.6 | 3118 | 4 | US-09-579-181-1 |
| 12 | 191.5 | 4.5 | 894 | 3 | US-08-362-525-22 |
| 13 | 191.5 | 4.5 | 894 | 3 | US-08-971-692-15 |
| 14 | 184.5 | 4.3 | 1537 | 1 | US-08-325-267A-2 |
| 15 | 178 | 4.2 | 1721 | 3 | US-08-700-651-5 |
| 16 | 178 | 4.2 | 1721 | 3 | US-08-928-361B-6 |
| 17 | 175.5 | 4.1 | 3969 | 4 | US-08-061-376-5 |
| 18 | 172 | 4.0 | 750 | 4 | US-09-165-239A-4 |
| 19 | 168 | 3.9 | 1837 | 3 | US-08-928-361B-5 |
| 20 | 166 | 3.9 | 805 | 4 | US-09-103-429A-4 |
| 21 | 164.5 | 3.9 | 752 | 1 | US-08-244-189-2 |
| 22 | 164 | 3.8 | 903 | 2 | US-08-853-310-2 |
| 23 | 160 | 3.8 | 1125 | 4 | US-09-513-783A-152 |
| 24 | 160 | 3.8 | 1610 | 4 | US-09-513-783A-22 |
| 25 | 158.5 | 3.7 | 786 | 4 | US-09-103-429A-3 |
| 26 | 157.5 | 3.7 | 941 | 4 | US-07-757-022B-14 |
| 27 | 157.5 | 3.7 | 1022 | 4 | US-07-757-022B-84 |

28 157.5 3.7 1038 4 US-07-757-022B-74 Sequence 74, Appl
29 157.5 3.7 1049 4 US-07-757-022B-58 Sequence 58, Appl
30 157.5 3.7 1140 4 US-07-757-022B-104 Sequence 104, App
31 157.5 3.7 1270 4 US-07-757-022B-44 Sequence 44, Appl
32 157.5 3.7 1311 4 US-07-757-022B-42 Sequence 42, Appl
33 157.5 3.7 1313 4 US-07-757-022B-142 Sequence 142, App
34 157.5 3.7 1314 4 US-07-757-022B-50 Sequence 50, Appl
35 157.5 3.7 1320 4 US-07-757-022B-46 Sequence 46, Appl
36 157.5 3.7 1320 4 US-07-757-022B-60 Sequence 60, Appl
37 157.5 3.7 1354 4 US-07-757-022B-48 Sequence 48, Appl
38 157.5 3.7 1361 4 US-07-757-022B-40 Sequence 40, Appl
39 157.5 3.7 1363 4 US-07-757-022B-52 Sequence 52, Appl
40 157.5 3.7 1404 4 US-07-757-022B-2 Sequence 2, Appl
41 157.5 3.7 1404 4 US-07-757-022B-62 Sequence 62, Appl
42 157 3.7 2843 1 US-07-741-940-2 Sequence 2, Appl
43 157 3.7 2843 1 US-08-289-548A-2 Sequence 2, Appl
44 157 3.7 2843 1 US-08-452-654-2 Sequence 2, Appl
45 157 3.7 2843 2 US-08-370-235A-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-725-012-2
; Sequence 2, Application US/08725012
; Patent No. 5710025
; GENERAL INFORMATION:
; APPLICANT: Dikstein, Rivka
; APPLICANT: Tili, Robert
; TITLE OF INVENTION: B-Cell Specific Transcription Factor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/725,012
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-725-012-2

Query Match 93.6%; Score 3990; DB 1; Length 801;

Best Local Similarity 99.9%; Pred. No. 9,6e-319;

Matches 800; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTLVTKVAPVSAPPKVSSGPRLPAPQIVAVKAPNTTTIQFPANLQLPPGTVLIKNSGGL 60

Db 1 GTLVTKVAPVSAPPKVSSGPRLPAPQIVAVKAPNTTTIQFPANLQLPPGTVLIKNSGGL 60

QY 61 MLVSPOOTVTRATTSNITSRPAPNPOTVKTCTVNSSOLIKKVVAVTPVKLAQIGT 120

Db 61 MLVSPQQTVAETTSNTTSRAVPANPQTVKICTVPSNSSQLIKKVAVTPVKLAQIGT 120
Qy 121 TVTTPKPSVQSVAVPTSVVTPGKPLNTVTTLKPSILGASSTPSNEPNKAENSA 180
Db 121 TVTTPKPSVQSVAVPTSVVTPGKPLNTVTTLKPSILGASSTPSNEPNKAENSA 180
Qy 191 VOINLSPMLNKKCNFLKWLAKLKLACSGSPGQNVKLVQELLDKIEAEFTPK 240
Db 181 VOINLSPMLNKKCNFLKWLAKLKLACSGSPGQNVKLVQELLDKIEAEFTPK 240
Qy 241 LYVELKSSPQHLVPLFKSVVALLRQLLPNSQSFQOCVQQTSSDMVATCTTIVTSPV 300
Db 241 LYVELKSSPQHLVPLFKSVVALLRQLLPNSQSFQOCVQQTSSDMVATCTTIVTSPV 300
Qy 301 VTTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTTA 360
Db 301 VTTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGVTVLHSGVPTAATGGTTA 360
Qy 361 GTGLLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTSLSPAVTFGETSGAAICLPVSKP 420
Db 361 GTGLLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTSLSPAVTFGETSGAAICLPVSKP 420
Qy 421 VVSCWDHICKPVICTPVOIKLAQPGVPLSQPAGIPTGSSSKQLFSLFHVVOQPSGGNEK 480
Db 421 VVSCWDHICKPVICTPVOIKLAQPGVPLSQPAGIPTGSSSKQLFSLFHVVOQPSGGNEK 480
Qy 481 QVTTISHSTLTITQKCGQKTPMVTIIPTSQFPASILKQITLPCNKITLSLQASPTQKNR 540
Db 481 QVTTISHSTLTITQKCGQKTPMVTIIPTSQFPASILKQITLPCNKITLSLQASPTQKNR 540
Qy 541 IKENVTSQFREDDDINDVTSMAGVNLNENACILATNSLSELVGTLIQSCCKDEFFLFIGALQ 600
Db 541 IKENVTSQFREDDDINDVTSMAGVNLNENACILATNSLSELVGTLIQSCCKDEFFLFIGALQ 600
Qy 601 KRILDIGKKHDTIENSDAVNLIQATQERLGLLEKLTATAQHRMTYKASENYILCSD 660
Db 601 KRILDIGKKHDTIENSDAVNLIQATQERLGLLEKLTATAQHRMTYKASENYILCSD 660
Qy 661 TRSKLFLEKLDQKQKRDLEEREMLLKAAKSRNKEDPQLRLKQAKELQLELAQI 720
Db 661 TRSKLFLEKLDQKQKRDLEEREMLLKAAKSRNKEDPQLRLKQAKELQLELAQI 720
Qy 721 QHRDANLTATAAIGPKRKRPLESGIEGLKNDLLASGTSSTLTATQOLHRPRITRCLRLDI 780
Db 721 QHRDANLTATAAIGPKRKRPLESGIEGLKNDLLASGTSSTLTATQOLHRPRITRCLRLDI 780
Qy 781 FCMQEREMKYSRALLYALLK 801
Db 781 FCMQEREMKYSRALLYALLK 801

RESULT 2
US-08-188-582-16
; Sequence 16, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Slegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,582
FILING DATE: 28-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 737 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-582-16

Query Match 30.7%; Score 1307.5; DB 1; Length 737;
Best Local Similarity 40.0%; Pred. No. 1.3e-98;
Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;

Qy 13 PKPVSSG-----PRLPAQIVAVKAPNTTTIOFFANLQLPPTGTVLTKSNSGPLM 61
Db 68 PTATTSIGIRATLTPVLAQLPQP-----PQNTPIQ---NFQLPPGMVLRSENGQLL 118
Qy 62 LVSPQQTVTR-----AETTSNITSRPAVPANPQTVKICTVPSNSSQLIKKVAVTPVKKL 115
Db 119 MI-POQALAOQAOAHQAPQPTMAPRPAPTTSAPPVQISTVOAGPTIAR-QVTP---- 172
Qy 116 AQIGTVTVTPKPSVQSVAVPTSVVTPGKPLNTVT--TLKPSSILGAS--TPS 168
Db 173 ---TTIHKV---SQQTTVQPSATLQSPGVQVQLVGGAAQTASLTATAVGTGPQ 225
Qy 169 NE-PNLKAENSAVQINLSPTMLENVKCKNFKLMLKILACSGSQSPGQNVKLVQEL 227
Db 226 RTVPQATTTSSAATE-----TMENVKCKNFKLMLKILACSGSQSTETAANVKELVQNL 279
Qy 228 LDKAIEAEFTPKLYVELKSSPQHLVPLFKSVVALLRQLLPNSQSFQOCVQVQ----TS 283
Db 280 LDGKIEAEFTSRLLYRELNSSPQPYLVPLFKRSLPALRLQLTPDSSAAFTQQSQQQPPPTTS 339
Qy 284 SDMVIATCTTIVTSPVVTTVSSSQSEKSIIVSGATAPRTVSVQTLNPLAGPVGAKAGV 343
Db 340 Q-----ATTALTAVLSSSVORTAGTAAVTTSALQPPVLSL----- 376
Qy 344 VTLHSGVPTAATGGTAGTGLLQTSKPLVTSVANTVTVSLOPEKPVVSGTAVTSLSPAV 403
Db 377 -----TQPTQVGVGKGOPTPLVIQ-----QPKP-----CALIRPPQV 410
Qy 404 TFGETSGAAICLPVSKPVVSVSCWDHICKRPVICTPVOIKLAQPGVPLSQPAGIPTSSSKQ 463
Db 411 TLTQT-----PMVALRQPH-NRIMLTTPQOIQL----- 437
Qy 464 LFLSFHVVOQPSGGNEKQVTTISHSSTLTITQKCGQKTPMVTIIPTSOFFPASPILKQITL 523
Db 438 -----NLPQPVVVKPAVL 451
Qy 524 PGNKIL-----SIQASPTQKNRIKENVTSCFREDDDINDVTSMAGVNLNENACILATNSL 580
Db 452 PGTKALSAVSAQAAAAQKNLKEPGGSGFRDDDDINDVASMAGVNLSEESARILATNSL 511
Qy 581 VGTLIQSCCKDEFFLFIGALQKRLIDIGKKHDTIENSDAVNLIQATQERLGLLEKLTATA 640
Db 512 VGTLTRCKDETFLLQAPLQRRILEIGKKHGITELHPDQVSVYVSHATQORLQNLVEKISE 571

QY 641 IAQHRMTTKASENYILCSOTRSQALFKLEKLDQLEKQKDLBEREMLLKAASRSNKEDP 700
DB 572 TAQKNFYSYKDDRYEQASDVRAQLKFFQDQIEKQKDEQEREILMRAAKSRSRQEDP 631
QY 701 EQLRLKQAKELQOQLAQIHRDANLATAAIGPKRKRPLE-----SGIEGLKDNLLAS 755
DB 632 EQLRLKQAKEMQOQELAQMRQDANLATAAIGPKRKRKVDGPGSGAEGSGPGSVVP 691
QY 756 GTSSTATKQLHRPRITRCLRDILFCMEQEREMKYSRALLYLALK 801
DB 692 GSSGVGTPTQRTQRTITRNLRLDILFCLENERETSHLLLYKAFK 737

RESULT 3

US-08-646-715-16
; Sequence 16, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-646-715-16

Query Match 30.7%; Score 1307.5; DB 1; Length 737;
Best Local Similarity 40.0%; Pred. No. 1.3e-98;
Matches 330; Conservative 103; Mismatches 200; Indels 193; Gaps 22;

QY 13 PKVSSG-----PRLPAPQIVAVKAPNTTTIQFPANLQPLPGTVLIKSNSGPLM 61
DB 68 PTATTSIGIRATLTPTVLAPRLPQP-----PQNPTNIQ---NFQLPPGMVLVRENGQLL 118

QY 62 LVSPQQTVTR-----AETTSNITSRAVPANPOTVKICTVPSNSSOLIKKVAVTPVKKL 115
DB 119 MI-POQALAQMAQAHAQPOTTMAPRAPTPTAPPVQISTVAPGPTIAR-QVTP----- 172
QY 116 AQIGTAVTVPKPSVQSVAVPTSVVTPGKPLNTVT--TLKPSSSLGASS-----TPS 168
DB 173 ----TWIIKQV---SOAQTTVQPSATLQSRFGVQPOLVLCGAATASLGATATVQTGTPQ 225
QY 169 NE-PNLKAENSAAVQINLSPTMLENVKCKNFAMLIKACSGSQSPSEMGNVKKLYEQL 227
DB 226 RTVPGATTSSAATE-----TMENVKCKNFSTLTIKLASSGQSTETAAANVKELVONL 279
QY 228 LDKAIEAEETRKLYVELKSSPOPHLVPLKSSVVALROLLPNSQSFIQCVQ-----TS 283
DB 280 LDGKIEAEDFTSRLYRELNSSPOPLVPLFKRSLPALRLTPDSAAFIQSSQOQPPPTS 339
QY 284 SDMVIATCTVTTSPVTTTSSSQSEKIIIVSGATAPRTSVQTLNPLAGPVGAKGV 343
DB 340 Q-----ATTALTAVVLSSSVORTAGKTAATVTSALQPPVLSL----- 376
QY 344 VTLHVSPTAATGGTAGTGLQTSKPLVTSVANTVTVTSLOPEKPVVSGTAVTSLPAV 403
DB 377 -----TQPTQGVGKGQOPTPLVIQ-----QPKP-----GALIRPPQV 410
QY 404 TFCETSGAAICLPSVVPVSCWDHICKPVIGTPVQIKLAOPGVLSPQAGIPTGSSSKQ 463
DB 411 TLTOT-----PMVALRQPH-NRIMLTTPQOQL----- 437
QY 464 LFLSHVVOQPSGGNEKQVTTISHSSTLTIQCKGQKTMPTVNTIPTSQFPPASTLKQITL 523
DB 438 -----NPLQPVVVKPAVL 451
QY 524 PGNKIL---SLQASPTQKNRIKENVTSCFRDEDDINDVTSMAGVNLNEENACILATNSEL 580
DB 452 PGTKALSAVSAQAAAKNKLKEPGGSGFRDDDDINDVASWAGVNLSEESARILATNSEL 511
QY 581 VGTLIOSCKDEPFLFCALQKRIIDIGKKHIDITELNSDANLISOATQERLGLLEKLT 640
DB 512 VGTLTRSCCKDETFLLQAPLQRRILEIGKKHIGITELHPDVVSYVSHATQORLQNLVEKISE 571
QY 641 IAQHRMTTKASENYILCSOTRSQALFKLEKLDQLEKQKDLBEREMLLKAASRSNKEDP 700
DB 572 TAQKNFYSYKDDRYEQASDVRAQLKFFQDQIEKQKDEQEREILMRAAKSRSRQEDP 631
QY 701 EQLRLKQAKELQOQLAQIHRDANLATAAIGPKRKRPLE-----SGIEGLKDNLLAS 755
DB 632 EQLRLKQAKEMQOQELAQMRQDANLATAAIGPKRKRKVDGPGSGAEGSGPGSVVP 691
QY 756 GTSSTATKQLHRPRITRCLRDILFCMEQEREMKYSRALLYLALK 801
DB 692 GSSGVGTPTQRTQRTITRNLRLDILFCLENERETSHLLLYKAFK 737

RESULT 4

US-08-188-582-2
; Sequence 2, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco


```

Matches 256; Conservative 126; Mismatches 308; Indels 178; Gaps 26;
QY 25 PQIVAVKAPNTTIOFPAQLPGLTGLKSNGLML-----VSPQOTVTRAETTS 76
Db 133 POSPSTLTSTLNTGTPA-----LLVKTGDFQLLRVGTGTPPTVGTQITNTSNN 184
QY 77 NITSRPVAPNPQVVKICTVPSNSSQ-----LIKKVAVTPVKKLAQIGTIVTVP 127
Db 185 NTTSTNHPPTTQ-IRLOTVPAAASMTNTATSNIVNSVASSGYANSSQPHLTQLNAQ 243
QY 128 KPSSVQSVAVPTSVVTPGKPLNVTTLKPSLSGASTPSNEPNLKAENSAVAOINLSP 187
Db 244 APQLPQITQITIPAQSOQOQVNNVSSAGGTAVSSTTA-----ATT 287
QY 188 TMLSNVR-KCNFLAMLIKACSSQSPQEMCONVKLVQLLDAKIEAEETRLYLVELK 246
Db 288 TOQNTKEKCKRKLANLIEL--STREPKPVEKNVTRTLQIELVNAVPEEFCDLRELLN 345
QY 247 SSPQHLVPLFLKKSVALRQL-----LPSQSFIQ--- 276
Db 346 ASPQCLIGFLKKSLLPRLRLALYTKELVIEGKPPQHVGLCLAGLSQQLPQIAQIRPIG 405
QY 277 ----OCVOQTSDDMVIACTTFTVTTPSPVTTTVSSSSEKSIIVSGATAPRTVS---VQT 329
Db 406 PSQTTTIGQTOVRMI--TPNALGTTPRTTIGHTTISKOPPN---IRLPTAPRLVNTGGIRT 460
QY 330 LNPLAGPVGAKGVVTLHSVGPATAGTGTAGTGLLOTSKPLVTSVANTVTVLSLOPEKP 389
Db 461 QIP-SLOVPGQANVQIR--GPQHAQLORTGSVQIRATRP-----PNSVPTAN----- 506
QY 390 VVSGTAVTSLPVPFTGETSGAAICLPSKPVVSCFWDHICKPVIGTPVQIKLAQPGPVL 449
Db 507 -----KLTAVKVQGTQIKAI-TPSLHP-----PSLAAISGGP-----PPTPTL 543
QY 450 SOPAGITPSSSK---QLFSLFHVQVQPSGGNEKQVTTISHS----- 489
Db 544 SVLSTLNSASTTILPISLPTVHLPPALRAREQOMQNSLNHNSHFDKLVKIKAPSLHP 603
QY 490 -----TLTIQCKGOKMPVNTIIPTSQFPASILKQITLPGNKILSLQ-----AS 534
Db 604 PHMERINASLTPICAKTM-----ARPPPAINKAIGKKRDRAMENDAKLNTSSGAA 654
QY 535 PTKNRKIKENTVTSFREDDINDVTSMAVNLNEENACILATNSLSELVGLTIQSKDEPFL 594
Db 655 SAANSFQOQSSMSMYGDDINDVAAMGVNLAESORILCC-TENIGTQIRSKDEVEL 713
QY 595 FIGALQKRLDIGKHHITELNSDAVNLISOATQERLGLLEKLTATAOHRMTYTKASEN 654
Db 714 NLPSLQARIRAITSEAGLDEPSQDVAVLISHACQERLKNIVEKLAVIAEHRIDVILKDP 773
QY 655 YILCSDTRSQLKFLKLDLEKQKRDLEEREMLLKAASRKNKEDPQLRLKQKAKELQ 714
Db 774 YEPADKVRGQIKFLELDKAPQKHELEEREMLLRAAKSRVDEPQAKMKARAKEMOR 833
QY 715 LELAQIQRDANLATAAIGPRKRPLE-----SGIEGLKDNLLASGTSLSLTATKQLHRP 769
Db 834 AEMELQRDANLALQAIGPRKKLKDGETVSSGASGGCVLLSSGCSAPTTL-----RP 889
QY 770 RITRICLRDLFCMEQEREMKYSRALLYLALK 801
Db 890 RIKRVNLRDLFMEQEREFRCRSMLEKTYLK 921

RESULT 6
US-08-046-585-5
; Sequence 5, Application us/08046585
; Patent No. 5453362
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR

```

```

; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/046,585
; APPLICATION NUMBER: US/08/046,585
; FILING DATE: 12-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-046-585-5

Query Match 5.6%; Score 237; DB 1; Length 2035;
Best Local Similarity 22.6%; Pred. No. 4.2e-10;
Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;
QY 10 VSAPPKVSSGPRLPAPQIVAVKAPNTTIIQPPANQLPPGTVLTKSNLMLVSPQOVT 69
Db 537 IGSSPSSGMAALAAAAATOKIIPSSA---PTVLSVPAGTIVTKT-----MAVTPGTTT 588
QY 70 TRAEITSNITSRPVPANPOTVKTCTVPNSSOLIKKVVAVTPVKKLAQIGTIV----- 122
Db 589 LPA--TVKVASPVVWVSNPAT-----RMLKTA-----AQGVTSVSATNTS 628
QY 123 ---VTVPKPSSVQSVAVPTSVVTPGKPLNTVTLK-PSSL-GASSTPSNEPNL----- 173
Db 629 TRPIITVHKSGTV-TVAQQAQVTVTVVGGVTKTITLVKSPISVPGSALINLGVMSV 687
QY 174 --KAENSAAVOINISPTMLENVKCKNPL--AMLIKACSSQSP-----EMGNVK 221
Db 688 QTKPVQTSVAVTQASTGPTVQIQTGKPLPAGTILKLVTSADGKPTTIIITTTQASGATK 747
QY 222 KLVQLLDAKIEAEETRLKLVKSSPQPHLVPLFKSVVALLQLLPNSQSFQOQVQ 281
Db 748 PTILGI-----SSVSPSTT---RPGTTIITIKTIPMSAIIITQAGATG 785
QY 282 TSSDMVIATCTTIVTT-----SPVVT-----TTVSSSQSEKSIIVSGA-----TA 321
Db 786 VTSSPGIKSPITIIITTKVMTSGTGAPAKIITAVPKIATGCGQGVTVQVVLKAGAPQGTI 845
QY 322 PRT-----VSVOTLNPAGPVGAK--AGVTLHVSVPATAAGTGTAGTGLLOTS 368
Db 846 LRTVPMGGVRLVTPVTVSAVKPAVTVLVKGTGTGTTIGTVGTST--SLAGAGGHSTS 903
QY 369 KPLVTSVA--NTVTVTSLOPEKPVVSGTAVTSLPVTFTGETSGAAICLPSVKPVVSCW 426
Db 904 ASLATPITTLGTIATLSSQ-----VINPTAITVSAQVTLTAAGGLTPTITMQPV----- 954
QY 427 DHICKPVIGTPVQIKL-AQCPGVLSQPA-GIPTSSSKQLFLSFLWVQOQPSGGNEKQVTT 484
Db 955 -----SQPTQVTLITAPSGVEAQPVHDLVPS-----ILASP-----TT 987

```

QY 485 ISHSTLTITQKCGQKMPVNTIPTISQFPASI-----LKQITLPGNKILSLQASPTQ 537
Db 988 EOPTATVTIADSGQGVQGTVTLVCSNPPCETHETGTTATTIVAN--LGGHPQPTQ 1045
QY 538 KNRIKENVTSQFDEDDINDVTSMAGVNLEENACILATNSLGVHLIOCKDEP 592
Db 1046 VQFV-----CDRQEAASLVSTVG-----QQN-----GSVVRVCSNPP 1079

RESULT 7
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,703
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-2/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-393-703-5

Query Match 5.6%; Score 237; DB 1; Length 2035;
Best Local Similarity 22.6%; Pred. No. 4.2e-10;
Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;

QY 10 VSAPPKVGSLPAPQIVAVKAPNTTTIQFPANLQLPPTGVLIKSNGLPLVSPQQTQ 69
Db 537 ICSSQMSGMAALAAAAATQKIPSSA---PTVLSVPAGTTIVKT-----NAVTPGTTT 588
QY 70 TRAEITSNITSRPVNPANPOTVKICTVPNSSSOLIKKAVTPVKKLAQIGTTV-----122
Db 589 LPA--TVKVASSPVMVSNPAT-----RMLKTA-----AQVTSVSSATNTS 628
QY 123 ---VTVPKPSSVQSVAVPTSVVTPGKPLNTVTLK--PSSL-GASSTPSNEPNL----173
Db 629 TRPIITVHKSGTV--TAAQAAQVTTVVGVTITVLKSPISVPGSGSALISLGNKVMVW 687
QY 174 --KAENSAVQNLSPMLNENKCKNFL--AMLKILACSGSQSP-----EMQNVK 221
Db 688 QTKPVQTSVTCQAGTGPVTVIIQTKGKPLPAGTILKLVTSADGKPTTITTTQASGAGTK 747

QY 222 KLVEQLLDAKIEAEFERKLYVELKSSPOPHLVFELKSVVALROLPLNSQFIQCVCQ 281
Db 748 PILGI-----SSVSPST---KPGTTTIKIPMSAITQAGATG 785
QY 282 TSDMVIATCTTIVTT-----SPVVT-----TVSSSQSEKSIIVSGA-----TA 321
Db 786 VTSSPGIKSPITIIITKVTMTSGTAPAKIITAVPKIATGHGQGVTVVVLKGAQGPQGTI 845
QY 322 PRT-----VSVQTLNPLAGVPGAK--AGVTLHSVGPATAGTTAGTGLQTS 368
Db 846 LRTVPMGGVRLVTPVTSVAVKPAVTTLVVKGTTGTTGLTGTVTGTVST--SLAGAGHSTS 903
QY 369 KPLVTSVA--NTVTVSLQPEKPVVSGTAVTLSLPAVTEGETSGAAICLPSVKPVVSEFW 426
Db 904 ASLATPITTLGTIATLSSQ-----VINPTAITVSAAGTTTAAAGLLTPTITMPQV-----954
QY 427 DHICKRVICTPVQIKL-AQGPVLVSQA-GIPTGSSSKOLFSLFHVVOQPSGSGNEKQVTT 484
Db 955 -----SQPTQVTLITAPSGVEAQPVHDLPS-----ILASP-----TT 987
QY 485 ISHSTLTITQKCGQKMPVNTIPTISQFPASI-----LKQITLPGNKILSLQASPTQ 537
Db 988 EOPTATVTIADSGQGVQGTVTLVCSNPPCETHETGTTATTIVAN--LGGHPQPTQ 1045
QY 538 KNRIKENVTSQFDEDDINDVTSMAGVNLEENACILATNSLGVHLIOCKDEP 592
Db 1046 VQFV-----CDRQEAASLVSTVG-----QQN-----GSVVRVCSNPP 1079

RESULT 8
PCT-US93-11721-5
; Sequence 5, Application PC/TUS9311721
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11721
; FILING DATE: 03-DEC-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-57503-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-11721-5

Query Match 5.6%; Score 237; DB 5; Length 2035;
Best Local Similarity 22.6%; Pred. No. 4.2e-10;
Matches 148; Conservative 84; Mismatches 239; Indels 184; Gaps 32;

QY 10 VSAPPKSSGRLPAPQIVAVKAPNTTTIOPANLQLPPTGTVLKSNGLMLVSPQOQTV 69
DB 537 IGSSPQMSGMAALAAAAATOKIPSSA---PTVLSVPAGTTIVKT-----NAVTPGTTT 588
QY 70 TRAEITSNITSRPAPVAPNPQVKTCTVPNSSSOLIKKVVAVTPVKLAQIGTVV----- 122
DB 589 LPA---TVKVASPVNVMSPAT-----RMLKTA-----AOVGTSVSAINTS 628
QY 123 ---VTVTPKSSVOSVAVPTSVWTPGKPLNTVTLK-PSSL-GASSTPSNEPNL----- 173
DB 629 TRPIITHKSGTV-TVAQAAQVTVTVGGVTKTTLKSPISVPFGGSALISNLKVKVMSV 687
QY 174 --KAENSAVQINISPTMLENVKCKNPL--AMLIKACSSQSP-----EMGNQVK 221
DB 688 QTKPVQTSVAVTQASTGPVTQIQTKGPLPAGTILKLVTSADGKPTTIIITTTQASGAGTK 747
QY 222 KLVQLLDKAEAEFFKRLVLELSSPOPHLVFLKSVVALLRQLLPNSQSFIOQCVQ 281
DB 748 PTILGI-----SSVSPSTT---KPGTTTIKTIPMSAIIITQAGATG 785
QY 282 TSDMWIATCTTTVT-----SPVVT-----TVSSSQSEKSIIVSGA-----TA 321
DB 786 VTSFPGIKSPIITTTKMTSGTCAPAKIITAVPKIATGCGQGVTVVLLKAGAPQPGTI 845
QY 322 PRT-----VSVOTLNPLAGPVGAK--AGVTLHSGPTAAGTTAGTGLLQTS 368
DB 846 LRTVPMGVRVLTVPVTSVAVKPAVTTLVVKGTTGTVTLGTGTGTVST--SLAGAGHSTS 903
QY 369 KPLVTSVA--NTVTVSLQPKPVVSGTAVTSLPVPVFTGTSAAICLPSVKPVVSPCW 426
DB 904 ASLATPITTLGTIATLSQ-----VINPATAVSAQTTTLTAAGGLTTPTITMQPV----- 954
QY 427 DHICKPVIGTPVQIKL-AQPGVLSOPA-GIPTSSSKQLFLHVVOOPSSGNEKQVTT 484
DB 955 -----SQTQVLTAPSGVCAQVHDLPVS-----ILASP-----TT 987
QY 485 ISHSTLTIQCGOKTMPVNTIIPTSQPPASI-----LKQITLPGNKILSLQASPTQ 537
DB 988 EQPATVTIADSGQGVQPGVTLVCSNPPCETHETGTNTATTIVVAN--LGHPQPTQ 1045
QY 538 KNRKENVTSFREDDDINDVTSMAGVNLNEENACILATNSELVGTLIQSCKDEP 592
DB 1046 VQFV-----CDROEAAASLTSTVG-----QON-----GSVVRVCNPP 1079

RESULT 9

US-08-325-267A-4
; Sequence 4, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENITILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

Query Match 4.6%; Score 198; DB 1; Length 862;
Best Local Similarity 19.1%; Pred. No. 1.7e-07;
Matches 145; Conservative 110; Mismatches 278; Indels 228; Gaps 29;

QY 11 SAPPKVSSGRLPAPQIVAVKAPNTTTIQF---PANLQLPPTGTVLKSNGL----- 60
DB 199 SLPPNIEGTVMYAGYYPMKVYSNVSWCTLPISVTLPDGTTVSDDFEGVTVSFDL 258
QY 61 -----MLVSPQOVTTRAET-----TSNITSRPAPVAPNPQVKTCTVPNS 99
DB 259 SQSNCTVPDPSNAYATTTTTPWTGFTSTSTMTTGTNGVPTD-ETVIVIRTTT 317
QY 100 SSOLIKKVVAVTPVKAQIGTVTVTPKSSVOSVAVPTSVWTPGKPLN-TVTLK- 157
DB 318 AS-----TIIITTEPWTGFTSTSTMTTGTNGQPTDEIVIVIRT 359
QY 158 PSSLGASSTPSNEP-----NLKAENSAVQINISPTMLENVKCKNPLAMLIKACSGS 211
DB 360 PTEGLVTT-TTEPWTGFTSTSTMTSTVTCNGLPTD-ETVIVVKT-PTTAISSSLSS 416
QY 212 QSPMGQNVKLVQLLDKAEAEFFKRLVLELSSPOPHLVFLKSVVALLRQLLPNS 271
DB 417 SSGQITSSI-----TSSRPITTFPYS-----NG 440
QY 272 QSFIOQCVQOQTSMDVIATCTTTV-TTSPVVTTVSS-----SQSEKSIIV----- 316
DB 441 TSVI-----SSSVISSVTSSLTFTSPVSSSVISSSTTSIFSEKSSKSVIPTSS 493
QY 317 --SGATAPRTVSVQTLNPLAGPVGAKGVVTLHSGVPTAAGTTAGTGLLQTKPLVTS 374
DB 494 STSGSSESETSS-----AGSVSSSPSISSEKSSKSPYSS-----SSLPLVTS 535
QY 375 VANTVTVSLQPEKPVVSGTAVTSLPVAFTGETSGAAICLPSVKPVVSVFCDHIC----- 430
DB 536 ATTSQETAS-----SLPATTTKTSEQTTLV-----TVTSCESHVCTESI 575
QY 431 KPVITGPVQIKLAQPGPVLSQAPAGIPTGSSSKQLFLSHVHVQVQSGGKQVTTISHST 490
DB 576 SPAIVSTATVTV-----SGVTTEYTWCPISITETTKQKGTETQTTKQTV 625
QY 491 LTIQK-----CQKTMVNTI-----IPTSQFPPASILKQITLPGNK 527
DB 626 VTISCESDVCSKTASPAIVSTSTATINGVTTEYTWCPISITETTSRQOQTLVTVTSCEG 685
QY 528 ILSQASP-----TQKNRIKENVT-----SCFREDDINDVTSMAGVNLNEENACILAT 576

Db 686 VCSETASPAIVSTATATVNDVTVYTPWRPQTANEESVSKMNSATGETTTNTLAAETTT 745
QY 577 NSELVGTLIQSCKDEPFLFIGALQKR-----ILDIGKKHDITELN 616
Db 746 NTVAAETITNGAETKTVVTSLSRHAETQATSDVIGHSSSVSVSETGNTKSLT 805
QY 617 SDAVNLISQATQRLGLLEKLAIAHQHRTWYKASENVIL 657
Db 806 SSGLTMSQOPRSTPASSVMVGST-ASLEISTVAGSANSLL 845
RESULT 10
US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match 4.6%; Score 197; DB 4; Length 2972;
Best Local Similarity 23.7%; Pred. No. 1.5e-06;
Matches 140; Conservative 75; Mismatches 240; Indels 136; Gaps 32;
QY 2 TLVTKVAPVSAPPKVSQG-----PRLPAPQIVAVKAPNTTITQFPANLQL-- 46
Db 1063 TPVPLAPAPRPP--SSGLPAVLNPRPTLTGRLPTTLGTARAPMPTTLVRPLKLKLVH 1120
QY 47 --PPGTVLKNSGGLMLVSPQOTVTRAETTSNITSRPAVPANPQVVKICTVPSNSSQLI 104
Db 1121 SPSEFVSASAGNAPLTISPLHVP-----SSLPGPASPMP-----PIPNSSPLAS 1166
QY 105 KKVAVTPVKKLAQIGTVTVTVTPKPSVQSVAVPTSV-VTVTPGKP--LNTVT-TLKPS 160
Db 1167 PVSSTVSFPLSSSLPISVPTLPAPASA-PLTIPISAPLTVSASGPAALLTSVTPPLAPVV 1225
QY 161 LGASSTPSNEPNLKAENSAVQINL--SPTMLENVKCKNFAMLIKACSGSQSPMGQ 218
Db 1226 PAAPGPSPQLQSGASPSASALTGLATAPSLSSQTPGHPLL-----LAPTSSHVPGLNS 1280
QY 219 NVKKLVEQ-LLDAKIEAEFEFRKLYVELKSSQPH-----LVPFLKKSVMALRQLLPNS 271
Db 1281 TVAPACSPVLVPSALASPF-----PSAPNPAQAASLLAPASSASQALATPLAPMA 1332
QY 272 QSFIOQCVOQTSDMVIATCTTTVTTSVVTVTVSSSQSEKSIIVSGAT----- 320
Db 1333 AP--QTAILAPSPAPPLAP-LPVLAPSPGAAPVLASSQTPVPVMAFSPSTPGTSLASAPV 1389
QY 321 -APRTV-----SVQTL--NPLAGPVGAKAGVTVTLHVSQPTAA--TGCT-----TACTGL 365
Db 1390 PAPTVPVLPASSTQTMPLPAPVPSPLPSPASTQTL-ALAPALAPTLGGSSPSQTLSLCTGNP 1448
QY 366 QTSKPLVTVANTVTVSLQPEKPVVSGTAVTILSL-PATVFGETSAAIC-LPSVKPVVS 423
Db 1449 QGFPP-----TQTLSTPASPSSVPTPAQTLSLAPGPPGLPTQTLSLAPAPPLAP---- 1497
QY 424 FCWDHICKPVGITPV-----QIKLAQPCPVLS---QPAGIPT-GSSSKQLFSLFH 469
Db 1498 -----ASPVGPAPAHITLAPASSASLLAPASVQTLTSLPAPVPTLGPAAQTALAP 1551

QY 470 V-VQOPSGGNEKQVTTISHSTLTIOKCGOKTMPVNTI--IPTSQFPASI 517
Db 1552 ASTQSPA-----SQASSLVWSASGAAPLPVTMWSRLPVSKDEPDTL 1592
RESULT 11
US-09-579-181-1
; Sequence 1, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3118
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-1

Query Match 4.6%; Score 197; DB 4; Length 3118;
Best Local Similarity 23.7%; Pred. No. 1.6e-06;
Matches 140; Conservative 75; Mismatches 240; Indels 136; Gaps 32;
QY 2 TLVTKVAPVSAPPKVSQG-----PRLPAPQIVAVKAPNTTITQFPANLQL-- 46
Db 1209 TPVPLAPAPRPP--SSGLPAVLNPRPTLTGRLPTTLGTARAPMPTTLVRPLKLKLVH 1266
QY 47 --PPGTVLKNSGGLMLVSPQOTVTRAETTSNITSRPAVPANPQVVKICTVPSNSSQLI 104
Db 1267 SPSEFVSASAGNAPLTISPLHVP-----SSLPGPASPMP-----PIPNSSPLAS 1312
QY 105 KKVAVTPVKKLAQIGTVTVTVTPKPSVQSVAVPTSV-VTVTPGKP--LNTVT-TLKPS 160
Db 1313 PVSSTVSFPLSSSLPISVPTLPAPASA-PLTIPISAPLTVSASGPAALLTSVTPPLAPVV 1371
QY 161 LGASSTPSNEPNLKAENSAVQINL--SPTMLENVKCKNFAMLIKACSGSQSPMGQ 218
Db 1372 PAAPGPSPQLQSGASPSASALTGLATAPSLSSQTPGHPLL-----LAPTSSHVPGLNS 1426
QY 219 NVKKLVEQ-LLDAKIEAEFEFRKLYVELKSSQPH-----LVPFLKKSVMALRQLLPNS 271
Db 1427 TVAPACSPVLVPSALASPF-----PSAPNPAQAASLLAPASSASQALATPLAPMA 1478
QY 272 QSFIOQCVOQTSDMVIATCTTTVTTSVVTVTVSSSQSEKSIIVSGAT----- 320
Db 1479 AP--QTAILAPSPAPPLAP-LPVLAPSPGAAPVLASSQTPVPVMAFSPSTPGTSLASAPV 1535
QY 321 -APRTV-----SVQTL--NPLAGPVGAKAGVTVTLHVSQPTAA--TGCT-----TACTGL 365
Db 1536 PAPTVPVLPASSTQTMPLPAPVPSPLPSPASTQTL-ALAPALAPTLGGSSPSQTLSLCTGNP 1594
QY 366 QTSKPLVTVANTVTVSLQPEKPVVSGTAVTILSL-PATVFGETSAAIC-LPSVKPVVS 423
Db 1595 QGFPP-----TQTLSTPASPSSVPTPAQTLSLAPGPPGLPTQTLSLAPAPPLAP---- 1643
QY 424 FCWDHICKPVGITPV-----QIKLAQPCPVLS---QPAGIPT-GSSSKQLFSLFH 469
Db 1644 -----ASPVGPAPAHITLAPASSASLLAPASVQTLTSLPAPVPTLGPAAQTALAP 1697
QY 470 V-VQOPSGGNEKQVTTISHSTLTIOKCGOKTMPVNTI--IPTSQFPASI 517
Db 1698 ASTQSPA-----SQASSLVWSASGAAPLPVTMWSRLPVSKDEPDTL 1738
RESULT 12
US-08-362-525-22

Sequence 22, Application US/08362525
Patent No. 6027910
GENERAL INFORMATION:
APPLICANT: KLAS, FRANCISCUS M.
APPLICANT: SCHREUDER, MAARTEN P.
APPLICANT: TOSCHKA, HOLSER Y.
APPLICANT: VERRIPS, CORNELIS T.
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
PROTEIN
TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
PROTEIN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBAY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,525
FILING DATE: 04-JAN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92202080.5
FILING DATE: 08-JUL-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92203899.7
FILING DATE: 14-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/01763
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 213289/T7020(V)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-525-22

Query Match 4.5%; Score 191.5; DB 3; Length 894;
Best Local Similarity 20.1%; Pred. No. 6.1e-07;
Matches 123; Conservative 87; Mismatches 232; Indels 171; Gaps 25;

QY 22 LPAPQIVAVKAPNTTTFQFPANLQLPCTVLIKNSGPMPLVSPQ-----TVTRA 73
Db 265 VPDPNSNVAVSTTTT-----EPWGTFTSTSTMTVTGTNGVPTDEVIVIRTP 315
QY 74 TTSNITSRPVP-----ANPQTVKICTVPSNSSLIKKVVAV---TPVKKLAQIGTVVTV 126
Db 316 TSEGLISTTTPETWGTFTSTSTMTVTGTNGVPTDEVIVIRTPSE---GLISTTTE 371
QY 127 PKSSVQSVAVPTSVWTVTPCKPLN-TVTTLK-PSSIGASSTPSNEP-----NLKAENS 178
Db 372 PWTGFTTSTMTVTGTNGVPTDEVIVIRTPSEGLVTT-TTEPWTGFTTSTSTMS 430
QY 179 AAVQINLSPTMLENVKCKNFLAMLIKACSGSPGNGVKKLVQQLDAKIEAFT 238
Db 431 TVTGTNGLPTD-ETVIVVKT-PTTAISSLSSSSSGGITSSI----- 470
QY 239 RKLVELKSSPQHPVFLKSVVALLPNSQSFQCCVQQTSSDVIATCTTV-TT 297

Db 471 -----TSSRPITTFYPS-----NGTSVI-----SSSVISSSVTSSTFTS 505
QY 298 SPVVTTVTSS-----SQSEKSIIV-----SGATAPRTVSVOTLAPLAGVGA 341
Db 506 SPVSSSVISSTTSTSTIFSESSKSVIPTSSSTSGSESETS-----AGSVSSSS 558
QY 342 GVTVLHVGTAAGTAGTGLLOTSKPLVTSVANTVTVSVLOPEKPVVSGTAVTSLP 401
Db 559 FISSESSKSPYSS-----SSLPLVTATTSOETAS-----SLP 592
QY 402 AVTFGETSGAAICLPVSKPVVFCWDHIC-----KPVIGTPVQIKLAQPGVLSQPA 457
Db 593 PATTKTSEQTLV-----TVSCSHVCTESISPAIVSTATVTV-----SGVTT 637
QY 458 GSSSKQLFSLPHVVQPSGGNEKQVTTISHSTLTQKCGQKTMVNTIITSOFPASI 517
Db 638 EYTWCPISSTETTQTKGTTEQTTTQTTVVTVTSSC-ESDVCSTKSPAIVSTSTAT 696
QY 518 LKQITLPGNKILSLQASPTQKNRIKENVTSFCRDEDDINDVTSMAGVNLNEACILATN 577
Db 697 INGVTTEYTWCPISSTESRQQTTLVTVTS-----ESGVCSETAS 737
QY 578 SELVGTLIQCKD 590
Db 738 PAIVSTATATVND 750

RESULT 13
US-08-971-692-15
Sequence 15, Application US/08971692
Patent No. 6114147
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Immobilized proteins with specific binding
TITLE OF INVENTION: capacities and their use in processes and products.
NUMBER OF SEQUENCES: 40
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,692
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 894 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-971-692-15

Query Match 4.5%; Score 191.5; DB 3; Length 894;
Best Local Similarity 20.1%; Pred. No. 6.1e-07;
Matches 123; Conservative 87; Mismatches 232; Indels 171; Gaps 25;

QY 22 LPAPQIVAVKAPNTTTFQFPANLQLPCTVLIKNSGPMPLVSPQ-----TVTRA 73
Db 265 VPDPNSNVAVSTTTT-----EPWGTFTSTSTMTVTGTNGVPTDEVIVIRTP 315
QY 74 TTSNITSRPVP-----ANPQTVKICTVPSNSSLIKKVVAV---TPVKKLAQIGTVVTV 126
Db 316 TSEGLISTTTPETWGTFTSTSTMTVTGTNGVPTDEVIVIRTPSE---GLISTTTE 371
QY 127 PKSSVQSVAVPTSVWTVTPCKPLN-TVTTLK-PSSIGASSTPSNEP-----NLKAENS 178
Db 372 PWTGFTTSTMTVTGTNGVPTDEVIVIRTPSEGLVTT-TTEPWTGFTTSTSTMS 430
QY 179 AAVQINLSPTMLENVKCKNFLAMLIKACSGSPGNGVKKLVQQLDAKIEAFT 238
Db 431 TVTGTNGLPTD-ETVIVVKT-PTTAISSLSSSSSGGITSSI----- 470

; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match 4.28; Score 178; DB 3; Length 1721;
Best Local Similarity 20.08; Pred. No. 2.2e-05;
Matches 137; Conservative 82; Mismatches 293; Indels 172; Gaps 29;

| | | | |
|----|-----|--|-----|
| Qy | 35 | TTTTQFPANLQLPPGTIVLIKSNGLMLVSPQQTIVTRAETTSNITSRPAVPANPQTVKIC | 94 |
| Db | 286 | TT | 345 |
| Qy | 95 | TVPNSSOLIKKVAIVPKLAIQITVTVPKPSSVQSVAVPTSVVTVTPGKPLNVT | 154 |
| Db | 346 | TT | 398 |
| Qy | 155 | TLKPSSLGASSTPS-----NEPNLKAENSA-----VOINLSPTMLEN-- | 192 |
| Db | 399 | TTATTTTTTSETESVIRPDEMCWLEKNGECEKAGATVGVIGDKGRIENGMAFTMIPND | 458 |
| Qy | 193 | -----VKKCNFLAMLKILACSGSPGQNGQNVKKLVQQLLDKIE--AEFTRKLY | 242 |
| Db | 459 | THVFRFKVDVGTIISVRCRKGAGKLEFP-----DRSLDFTIPPVAGHNSCII | 508 |
| Qy | 243 | VELKSSQPHLVPELKKSVVALROLLPNSQSF-----IQQCV-----QQTSSDMVIA | 289 |
| Db | 509 | VGVSQDGIHVSPYSGSDVSLISAPIQPSSELFNEVYCDCTAKYGAHSGYQTSADFVTT | 568 |
| Qy | 290 | TC---TTTIVTSPVTTTVSSQSEKSLIVSGATAPRTVSQTLNPLAGPVGAKAGVVT | 346 |
| Db | 569 | TTAKPTTTTGGAPGQPTTTTGSPSKP--TTTTTKATTTTTLNP-----IITT | 616 |
| Qy | 347 | HSVGPATAGTGTAGTGLLOTSTKPLVTSVANTVTVSLQPEKPVV-----SGTAVTLSL | 400 |
| Db | 617 | TTQKPTTTT--TTKVPG-----KP---PIATTTTL-----KPIVTTTAKATTTTTV | 661 |
| Qy | 401 | PAVTFGETSGAAICLPSPVVFSCWDHICKPVIG-----TPVQIKLAQPGPVLSQAGIP | 456 |
| Db | 662 | PTTT--TTTKRDEMTTTTPL-----PDIGDIEITPIE----- | 694 |
| Qy | 457 | TGSSSKOLFSLFHVQPSGGNEKQVTTISHSSTLTQKCGQKTMVNTIIPTSQFPAS | 516 |
| Db | 695 | -----KMLDKYTRMIYDYSNG-----LLDSNDEPIGSOAG-----IADTSNLPVQ | 738 |
| Qy | 517 | ILKQITLPGNKLISLOASPTQKNRI-----KENVTSCFDEDDIN--D | 557 |
| Db | 739 | THKSTGLPIDPWGLPDPKSGNLVHPYNOTMGLSVSYLAARNLTV---DDETGLP | 795 |
| Qy | 558 | VTSMAGVNLNEENACILATNSLVGLTIQCKDEFFLFIGALQKRILDKKHDITELNS | 617 |
| Db | 796 | IDTLTGYP--DPVSLIPFNPPE-IGELFDPISDE--IMNGTIAGIVSGISASESLLSQS | 850 |
| Qy | 618 | DAVNLIQATQERLUGLLEKLTAI | 641 |
| Db | 851 | ALIDPATMVMVGEFGLINPATGV | 874 |

Search completed: February 16, 2003, 22:02:26
Job time : 40.8929 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run On: February 16, 2003, 22:29:40 ; Search time 6475 Seconds
(without alignments)
11497.294 Million cell updates/sec

Title: US-09-763-909-1
Perfect score: 2558
Sequence: 1 gggacctgtgacaaagt.....ctttattacttcttacctat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_ov.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_mu.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|--------|---------------|--------|----|-------------|---------------------|
| 1 | 2534 | 99.1 | 2556 | 6 | I81227 | I81227 Sequence 1 |
| 2 | 2534 | 99.1 | 2556 | 9 | HSTAF11 | Y09321 H. sapiens T |
| 3 | 1145.2 | 44.8 | 182805 | 9 | AC017007 | AC017007 Homo sapi |
| C 4 | 580 | 22.7 | 119081 | 2 | AC121320 | AC121320 Homo sapi |
| C 5 | 580 | 22.7 | 169240 | 9 | AC022069 | AC022069 Homo sapi |
| 6 | 580 | 22.7 | 190782 | 2 | AP002752 | AP002752 Homo sapi |
| 7 | 338.2 | 13.2 | 3603 | 6 | I23466 | I23466 Sequence 15 |
| 8 | 338.2 | 13.2 | 3603 | 6 | I45759 | I45759 Sequence 15 |
| 9 | 338.2 | 13.2 | 4233 | 9 | HSU75308 | U75308 Human Ttp-a |
| 10 | 337.2 | 13.2 | 3252 | 9 | HSTAF113 | V11354 H. sapiens m |
| 11 | 325.8 | 12.7 | 2196 | 10 | AY038601 | AY038601 Mus muscu |
| C 12 | 294.6 | 11.5 | 174222 | 2 | AP001197 | AP001197 Homo sapi |
| 13 | 294.6 | 11.5 | 175553 | 2 | AP001096 | AP001096 Homo sapi |
| 14 | 294.6 | 11.5 | 182884 | 9 | AC007996 | AC007996 Homo sapi |
| C 15 | 294.6 | 11.5 | 190782 | 2 | AP002752 | AP002752 Homo sapi |
| C 16 | 148.6 | 5.8 | 333300 | 2 | AC125091 | AC125091 Mus muscu |
| 17 | 148 | 5.8 | 16903 | 2 | AC016839 | AC016839 Homo sapi |
| C 18 | 148 | 5.8 | 61682 | 2 | AC118057 | AC118057 Homo sapi |
| C 19 | 147 | 5.7 | 277892 | 2 | AC125105 | AC125105 Mus muscu |
| C 20 | 146.6 | 5.7 | 205466 | 2 | AC127767 | AC127767 Rattus no |
| 21 | 143.8 | 5.6 | 3535 | 3 | S63550 | S63550 transcripti |
| 22 | 142.2 | 5.6 | 3745 | 3 | AY069807 | AY069807 Drosophi |
| 23 | 142.2 | 5.6 | 4615 | 3 | DROTAFL110X | L06861 Drosophila |
| 24 | 142.2 | 5.6 | 4615 | 6 | I23459 | I23459 Sequence 1 |
| 25 | 142.2 | 5.6 | 4615 | 6 | I45752 | I45752 Sequence 1 |
| 26 | 137.6 | 5.4 | 205466 | 2 | AC127767 | AC127767 Rattus no |
| 27 | 124.4 | 4.9 | 62974 | 2 | AC118056 | AC118056 Homo sapi |
| 28 | 121.8 | 4.8 | 277892 | 2 | AC125105 | AC125105 Mus muscu |
| 29 | 112.4 | 4.4 | 158509 | 2 | AC114177 | AC114177 Rattus no |
| C 30 | 109.8 | 4.3 | 62974 | 2 | AC118056 | AC118056 Homo sapi |
| C 31 | 99 | 3.9 | 158509 | 2 | AC114177 | AC114177 Rattus no |
| C 32 | 98.4 | 3.8 | 191918 | 2 | AC102692 | AC102692 Mus muscu |
| C 33 | 94.6 | 3.7 | 121982 | 9 | HS1107C24 | AL109911 Human DNA |
| C 34 | 92 | 3.6 | 333300 | 2 | AC125091 | AC125091 Mus muscu |
| C 35 | 91.2 | 3.6 | 64035 | 2 | AC119227 | AC119227 Mus muscu |
| C 36 | 89.6 | 3.5 | 194426 | 2 | AL663067 | AL663067 Mus muscu |
| C 37 | 71 | 2.8 | 112277 | 2 | AC113239 | AC113239 Tetraodon |
| C 38 | 62.4 | 2.4 | 2147 | 9 | AK097744 | AK097744 Homo sapi |
| C 39 | 58.6 | 2.3 | 59865 | 2 | AC014422 | AC014422 Drosophi |
| C 40 | 58.6 | 2.3 | 168479 | 3 | AC093454 | AC093454 Drosophi |
| C 41 | 58.6 | 2.3 | 176056 | 3 | AC010066 | AC010066 Drosophi |
| C 42 | 58.6 | 2.3 | 283821 | 3 | AE003528 | AE003528 Drosophi |
| C 43 | 57.6 | 2.3 | 7218 | 6 | I66494 | I66494 Sequence 14 |
| C 44 | 56.8 | 2.2 | 126712 | 9 | AL137077 | AL137077 Human DNA |
| C 45 | 56.4 | 2.2 | 201470 | 2 | AC113649 | AC113649 Rattus no |

ALIGNMENTS

| | | | | | | |
|------------|----------|---|---------|-----|--------|-----------------|
| RESULT 1 | I81227 | I81227 | 2556 bp | DNA | linear | PAT 10-JUN-1998 |
| LOCUS | I81227 | Sequence 1 from patent US 5710025. | | | | |
| DEFINITION | I81227 | | | | | |
| ACCESSION | I81227 | | | | | |
| VERSION | I81227.1 | GI:3209517 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | | Unknown. | | | | |
| ORGANISM | | Unclassified. | | | | |
| REFERENCE | | 1 (bases 1 to 2556) | | | | |
| AUTHORS | | Dikstein,R. and Tjian,R. | | | | |
| TITLE | | Cell-type specific transcription factor | | | | |
| JOURNAL | | Patent: US 5710025-A 1 20-JAN-1998; | | | | |
| FEATURES | | Location/Qualifiers | | | | |

| | | | |
|----|------|---|------|
| QY | 961 | AGCACCAGAACTGTGTCACTGCGAAACTTTTGAACCCACTTGTGTCAGTGGGAGCAAA | 1021 |
| DB | 960 | AGCACCAGAACTGTGTCACTGCGAAACTTTGAACCCACTTGTGTCAGTGGGAGCAAA | 1019 |
| QY | 1021 | AGCTGGAGTTGTGACACTTCATTCTGTGGGCCCACTGCTGCAACAGGAGGACAACAGC | 1080 |
| DB | 1020 | AGCTGGAGTTGTGACACTTCATTCTGTGGGCCCACTGCTGCAACAGGAGGACAACAGC | 1079 |
| QY | 1081 | TGGAACCTGGTTTGCTTCAGACTTCAAACCACTTGTGACATCTGTGGCAACACAGTGCAC | 1140 |
| DB | 1080 | TGGAACCTGGTTTGCTTCAGACTTCAAACCACTTGTGACATCTGTGGCAACACAGTGCAC | 1139 |
| QY | 1141 | CACGCTCTCACTGCAACCTGAAAGCCAGTTGTCTCTGGAACAGCAGTAACACACTGTCCCT | 1200 |
| DB | 1140 | CACGCTCTCACTGCAACCTGAAAGCCAGTTGTCTCTGGAACAGCAGTAACACACTGTCCCT | 1199 |
| QY | 1201 | TCCAGCAGTAACCTTTTGGAGAAACTTCAGGTGCAGCTATTTGTCTTCATCTGTGAAACC | 1260 |
| DB | 1200 | TCCAGCAGTAACCTTTTGGAGAAACTTCAGGTGCAGCTATTTGTCTTCATCTGTGAAACC | 1259 |
| QY | 1261 | TGTTGTTTCCCTCTGCTGGGACACACATCTGCAAGCCTGTTATTGGACCTCCAGTTCAAAT | 1320 |
| DB | 1260 | TGTTGTTTCCCTCTGCTGGGACACACATCTGCAAGCCTGTTATTGGACCTCCAGTTCAAAT | 1319 |
| QY | 1321 | CAAACTTGCACGCGGGCCCTGTCCTTTCAACAACAGCTGGGATTCCAACAGGCAGTTC | 1380 |
| DB | 1320 | CAAACTTGCACGCGGGCCCTGTCCTTTCAACAACAGCTGGGATTCCAACAGGCAGTTC | 1379 |
| QY | 1381 | AAGCAAGCACTATTCATTTGTTTCACTGTGTGACAGCCTTCAGGAGGCAATGAAA | 1440 |
| DB | 1380 | AAGCAAGCACTATTCATTTGTTTCACTGTGTGACAGCCTTCAGGAGGCAATGAAA | 1439 |
| QY | 1441 | ACAAGTGACCACAATTCACATTCCTCAACATTGACCATTCAGAAATGTGCACAGAGAC | 1500 |
| DB | 1440 | ACAAGTGACCACAATTCACATTCCTCAACATTGACCATTCAGAAATGTGCACAGAGAC | 1499 |
| QY | 1501 | GATGCCAGTGAACACCATATACTACTAGTCAGATTTCCCTCCAGCTCCATCTTAAAGCA | 1560 |
| DB | 1500 | GATGCCAGTGAACACCATATACTACTAGTCAGATTTCCCTCCAGCTCCATCTTAAAGCA | 1559 |
| QY | 1561 | AATTACCTCTGCCTGGAAATAAAATTTCTGTCACTTCAAGCATCTCCTACTCAGAAAAATA | 1620 |
| DB | 1560 | AATTACCTCTGCCTGGAAATAAAATTTCTGTCACTTCAAGCATCTCCTACTCAGAAAAATA | 1618 |
| QY | 1621 | GAATAAAGAGATGTAAACATCATGCTCCGAGATGAGGATGACATCAATGATGACTT | 1680 |
| DB | 1619 | GAATAAAGAGATGTAAACATCATGCTCCGAGATGAGGATGACATCAATGATGACTT | 1678 |
| QY | 1681 | CTATGGCAGGGGTCAACCTTTAATGAAGAAATGCCGTGATCTTAGCAACAACACTCTGAAT | 1740 |
| DB | 1679 | CTATGGCAGGGGTCAACCTTTAATGAAGAAATGCCGTGATCTTAGCAACAACACTCTGAAT | 1738 |
| QY | 1741 | TGGTTGGACACACTATTCAGTCATGATAAGATGAACCAATTTCTTTTATTGGAGCTCTAC | 1800 |
| DB | 1739 | TGGTTGGACACACTATTCAGTCATGATAAGATGAACCAATTTCTTTTATTGGAGCTCTAC | 1798 |
| QY | 1801 | AAAAGAGAACTTAGACATTTGTAAGAACATGACATACAGAACTTAACCTCTGATGCTG | 1860 |
| DB | 1799 | AAAAGAGAACTTAGACATTTGTAAGAACATGACATACAGAACTTAACCTCTGATGCTG | 1858 |
| QY | 1861 | TGAACCTTGATCTCCCAAGCAACACAGGAACGACTACGAGGCCCTTCAGAAAACTGACTG | 1920 |
| DB | 1859 | TGAACCTTGATCTCCCAAGCAACACAGGAACGACTACGAGGCCCTTCAGAAAACTGACTG | 1918 |
| QY | 1921 | CAATTGCTCAGCATCGAATGACTTACAGGCAAGTGAAATTCATCTGTGTAGTG | 1980 |
| DB | 1919 | CAATTGCTCAGCATCGAATGACTTACAGGCAAGTGAAATTCATCTGTGTAGTG | 1978 |
| QY | 1981 | ATACCAGTCCACAGCTCAAAATTTCTTGAAGAACTGGATCAATTCGAGAGCAGAGAAAGG | 2040 |
| DB | 1979 | ATACCAGTCCACAGCTCAAAATTTCTTGAAGAACTGGATCAATTCGAGAGCAGAGAAAGG | 2038 |
| QY | 2041 | ATTTTGAAGAAAGAGAAATGTTACTTTAAGGCAGCCAAAGAGTCGTTCTTAATAAGAAAGATC | 2100 |

REFERENCE
AUTHORS
TITLE
JOURNAL

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 182805)
Waterston,R.H.
Direct Submission
Submitted (29-MAY-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 182805)
Waterston,R.H.
Direct Submission
Submitted (08-AUG-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 182805)
Waterston,R.H.
Direct Submission
Submitted (10-AUG-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 8, 2002 this sequence version replaced gi:21240754.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics

Center project name: H_NH0119H12

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-67F3; the clone sequenced
to the right is RP11-224E18. Actual start of this clone is at base
position 1 of RP11-119H12; actual end is at base position 182805 of
RP11-224E18.

Data from AC108042 was used to finish this clone, AC017007.

FEATURES source

1. 182805
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UL"
/map="4"
/clone="RP11-119H12"
/clone_lib="RPCI-11"

repeat_region
1. 57
/rpt_family="Alu"
58. 344
/rpt_family="Alu"
345. 501
/rpt_family="Alu"
502. 627
/rpt_family="L1"
638. 942
/rpt_family="Alu"
998. 1150
/rpt_family="Alu"
1193. 1350
/rpt_family="Achoho"
1353. 1440
/rpt_family="MIR"
1443. 1734
/rpt_family="Alu"
1996. 2126
/rpt_family="Alu"
2133. 2441
/rpt_family="Alu"
2847. 2885
/rpt_family="(CAATT)n"
2966. 3270
/rpt_family="Alu"
3312. 3352
/rpt_family="(TA)n"
3877. 4024
/rpt_family="L1"
4299. 4557
/rpt_family="Alu"
4873. 5162
/rpt_family="L1"
5339. 5367
/rpt_family="AT-rich"
5377. 5677
/rpt_family="Alu"
5679. 5759
/rpt_family="MIR"
6137. 6176
/rpt_family="(TGG)n"
6206. 6427
/rpt_family="MIR"
6624. 7163
/rpt_family="MALR"
7234. 7323
/rpt_family="L2"
7324. 7629
/rpt_family="Alu"
7630. 7663
/rpt_family="L2"
8060. 8100
/rpt_family="(CAAA)n"
8527. 8820
/rpt_family="Alu"
9005. 9210
/rpt_family="Alu"
9426. 9618
/rpt_family="Alu"
9619. 9917
/rpt_family="Alu"
9918. 9928
/rpt_family="Alu"
9929. 10191
/rpt_family="Alu"
10192. 10324
/rpt_family="Alu"
10903. 11011
/rpt_family="Alu"
11015. 11043
/rpt_family="(CAAA)n"
11206. 11495
repeat_region

LOCUS AC022069 169240 bp DNA linear PRI 16-JUN-2002
 DEFINITION Homo sapiens, clone RP11-107K17, complete sequence.
 ACCESSION AC022069
 VERSION AC022069.6 GI:21429684
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 169240)
 TITLE Homo sapiens, clone RP11-107K17
 JOURNAL Unpublished
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 REFERENCE 2 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bada, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 DeAtrellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
 Norman, C.H., O'Connell, T., O'Donnell, P., Oliver, T.M., Peterson, R.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., DeAtrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (30-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., DeAtrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (30-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., DeAtrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,

Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 5 (bases 1 to 169240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., DeAtrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fato, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (16-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 16, 2002 this sequence version replaced gi:2126339.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L969
 Center clone name: 107_K_17

 Location/Qualifiers
 1..169240
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-107K17"
 /clone_lib="RPCI-11 Human Male BAC"
 complement(2..164)
 /rpt_family="L3"
 165..203
 /rpt_family="(TCA)n"
 complement(324..389)
 /rpt_family="L3"
 940..1236
 /rpt_family="AluSx"
 1657..1877
 /rpt_family="MIR"
 complement(2045..2391)
 /rpt_family="AluSx"
 complement(3134..3360)
 /rpt_family="L2"
 complement(3529..4480)

TITLE
JOURNAL

COMMENT

FEATURES
source

repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region

Quality coverage: 4.15x in Q20 bases; sum-of-contigs
NOTE: This is a 'working draft' sequence. It currently consists of
38 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

```
1 14910 contig of 14910 bp in length
20111 26140 contig of 11130 bp in length
36929 contig of 10689 bp in length
48911 contig of 11882 bp in length
57744 contig of 8733 bp in length
68839 contig of 10995 bp in length
77568 contig of 8629 bp in length
87657 contig of 9989 bp in length
93985 contig of 6228 bp in length
100964 contig of 6879 bp in length
107297 contig of 6233 bp in length
113948 contig of 6551 bp in length
120683 contig of 6635 bp in length
126925 contig of 6142 bp in length
130951 contig of 3926 bp in length
134620 contig of 3569 bp in length
138171 contig of 3451 bp in length
143486 contig of 5215 bp in length
148147 contig of 4950 bp in length
152197 contig of 3697 bp in length
155994 contig of 3352 bp in length
159446 contig of 3406 bp in length
162952 contig of 2504 bp in length
165556 contig of 2483 bp in length
168139 contig of 2483 bp in length
170861 contig of 2622 bp in length
173710 contig of 2749 bp in length
175823 contig of 2013 bp in length
177259 contig of 1336 bp in length
178684 contig of 1325 bp in length
180676 contig of 1892 bp in length
182266 contig of 1490 bp in length
184572 contig of 2206 bp in length
186271 contig of 1599 bp in length
188088 contig of 1717 bp in length
189189 contig of 1172 bp in length
190585 contig of 1125 bp in length
190782 contig of 97 bp in length.
NOTE: This is a 'working draft' sequence. It currently
consists of 38 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 14910: contig of 14910 bp in length
14911 15010: gap of 100 bp
15011 26140: contig of 11130 bp in length
26141 26240: gap of 100 bp
26241 36929: contig of 10689 bp in length
36930 37029: gap of 100 bp
37030 48911: contig of 11882 bp in length
48912 49011: gap of 100 bp
49012 57744: contig of 8733 bp in length
57745 57844: gap of 100 bp
57845 68839: contig of 10995 bp in length
68840 68939: gap of 100 bp
68940 77568: contig of 8629 bp in length
77569 77668: gap of 100 bp
77669 87657: contig of 9989 bp in length
87658 87757: gap of 100 bp
87758 93985: contig of 6228 bp in length
93986 94085: gap of 100 bp
```

```
* 94086 100964: contig of 6879 bp in length
* 100965 101064: gap of 100 bp
* 101065 107297: contig of 6233 bp in length
* 107298 107397: gap of 100 bp
* 107398 113948: contig of 6551 bp in length
* 113949 114048: gap of 100 bp
* 114049 120683: contig of 6635 bp in length
* 120684 120783: gap of 100 bp
* 120784 126925: contig of 6142 bp in length
* 126926 127025: gap of 100 bp
* 127026 130951: contig of 3926 bp in length
* 130952 131051: gap of 100 bp
* 131052 134620: contig of 3569 bp in length
* 134621 134720: gap of 100 bp
* 134721 138171: contig of 3451 bp in length
* 138172 138271: gap of 100 bp
* 138272 143486: contig of 5215 bp in length
* 143487 143586: gap of 100 bp
* 143587 148147: contig of 4950 bp in length
* 148148 148247: gap of 100 bp
* 148248 152197: contig of 3697 bp in length
* 152198 152297: gap of 100 bp
* 152298 155994: contig of 3352 bp in length
* 155995 156094: gap of 100 bp
* 156095 159446: contig of 3406 bp in length
* 159447 159546: gap of 100 bp
* 159547 162952: contig of 2504 bp in length
* 162953 163052: gap of 100 bp
* 163053 165556: contig of 2483 bp in length
* 165557 165656: gap of 100 bp
* 165657 168139: contig of 2483 bp in length
* 168140 168239: gap of 100 bp
* 168240 170861: contig of 2622 bp in length
* 170862 170961: gap of 100 bp
* 170962 173710: contig of 2749 bp in length
* 173711 173810: gap of 100 bp
* 173811 175823: contig of 2013 bp in length
* 175824 175923: gap of 100 bp
* 175924 177259: contig of 1336 bp in length
* 177260 177359: gap of 100 bp
* 177360 178684: contig of 1325 bp in length
* 178685 178784: gap of 100 bp
* 178785 180676: contig of 1892 bp in length
* 180677 180776: gap of 100 bp
* 180777 182266: contig of 1490 bp in length
* 182267 182366: gap of 100 bp
* 182367 184572: contig of 2206 bp in length
* 184573 184672: gap of 100 bp
* 184673 186271: contig of 1599 bp in length
* 186272 186371: gap of 100 bp
* 186372 188088: contig of 1717 bp in length
* 188089 188188: gap of 100 bp
* 188189 189313: contig of 1125 bp in length
* 189314 189413: gap of 100 bp
* 189414 190585: contig of 1172 bp in length
* 190586 190686: gap of 100 bp
* 190686 190782: contig of 97 bp in length.
FEATURES
    Location/Qualifiers
        1..190782
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="18"
        /map="18q11.2"
        /clone="Rp11-802C10"
        1..14910
        /note="assembly_fragment"
        15011..26140
        /note="assembly_fragment"
        26241..36929
        /note="assembly_fragment"
        37030..48911
        /note="assembly_fragment"
        49012..57744
        misc_feature
        misc_feature
        misc_feature
        misc_feature
        misc_feature
```

misc_feature /note="assembly_fragment"
57845..68839
misc_feature /note="assembly_fragment"
68940..77568
misc_feature /note="assembly_fragment"
77669..87657
misc_feature /note="assembly_fragment"
87758..93985
misc_feature /note="assembly_fragment clone_end:17 vector_side:left"
94086..100964
misc_feature /note="assembly_fragment"
101065..107297
misc_feature /note="assembly_fragment"
107398..113948

Query Match 22.7%; Score 580; DB 2; Length 190782;
Best Local Similarity 97.3%; Pred. No. 3.5e-141;
Matches 611; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY 769 TCCTAAGAAAAGCGTGGTGGCTTACGACAACTCTCGCTAACTCCAGAGCTTCATCCA 828
Db 166175 TTTCCAGAAAAGCGTGGTGGCTTACGACAACTCTCGCTAACTCCAGAGCTTCATCCA 166234

QY 829 GCAATGTTTCAGCAGACTTCTAGTGACATGGTCAATGCTACCTGTACTACAACAGTAAC 888
Db 166235 GCAATGTTTCAGCAGACTTCTAGTGACATGGTCAATGCTACCTGTACTACAACAGTAAC 166294

QY 889 AACTTCTCCTGTGGTGACAACTACAGTGTCTCAAGCAGCTCTGAAAAGTCAATTTATGT 948
Db 166295 AACTTCTCCTGTGGTGACAACTACAGTGTCTCAAGCAGCTCTGAAAAGTCAATTTATGT 166354

QY 949 TTTCTGGAGCAACAGACCCAGAACTGTGTAGTGCACAACTTTGACCCACTTGTGTGTCC 1008
Db 166355 TTTCTGGAGCAACAGACCCAGAACTGTGTAGTGCACAACTTTGACCCACTTGTGTGTCC 166414

QY 1009 AGTGGGAGCAAAAGCTGGAGTGTGTGACACTTCATTCTGTGGGCCAACTGTGTCAACAGG 1068
Db 166415 AGTGGGAGCAAAAGCTGGAGTGTGTGACACTTCATTCTGTGGGCCAACTGTGTCAACAGG 166474

QY 1069 AGGAACAACAGCTGGAACTGGTTTGTCTCAGACTTCAAAACCACTTGTGACATCTGTGGC 1128
Db 166475 AGGAACAACAGCTGGAACTGGTTTGTCTCAGACTTCAAAACCACTTGTGACATCTGTGGC 166534

QY 1129 AATACAGTGCACCGCTCTCACTGCAACCTGAAACCGAGTGTCTCTGGAACAGCAGT 1188
Db 166535 AATACAGTGCACCGCTCTCACTGCAACCTGAAACCGAGTGTCTCTGGAACAGCAGT 166594

QY 1189 AACACTGTCTCTCCAGCAGTAACTTTTGGAGAACTTTCAGGTGCAGCTATTGTCTTCC 1248
Db 166595 AACACTGTCTCTCCAGCAGTAACTTTTGGAGAACTTTCAGGTGCAGCTATTGTCTTCC 166654

QY 1249 ATCTGTGAAACCTGTGTGTCTCTCTCTGCTGGGACCACATCTG -CAAGCCCTGTTATTGGGA 1307
Db 166655 ATCTGTGAAACCTGTGTGTCTCTCTCTGCTGGGACCACATCTG -CAAGCCCTGTTATTGGGA 166713

QY 1308 CTCAGTTCAAATCAAACTTGGCCAGCCGGCCCTGTCTTTTCAACACAGCTGGGATTC 1367
Db 166714 CTCAGTTCAAATCAAACTTGGCCAGCCGGCCCTGTCTTTTCAACACAGCTGGGATTC 166773

QY 1368 CAACAGGCGAGTTCAGCAAGCAACTATT 1395
Db 166774 CAACAGGCGAGTTCAGCAAGCAACTATT 166801

RESULT 7
LOCUS 123466
DEFINITION Sequence 15 from patent US 5534410.
ACCESSION 123466
VERSION 123466.1 GI:1603336
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

Unclassified.
1 (bases 1 to 3603)
Tjian,R., Comai,L., Dynlacht,B.D., Hoey,T., Ruppert,S., Tanese,N.,
Wang,E. and Weinzierl,R.O.J.
TATA-binding protein associated factors drug screens
Patent: US 5534410-A 15 09-JUL-1996;
JOURNAL
FEATURES
source Location/Qualifiers
1..3603
BASE COUNT 969 a 961 c 830 g 843 t
ORIGIN

Query Match 13.2%; Score 338.2; DB 6; Length 3603;
Best Local Similarity 63.9%; Pred. No. 1.4e-77;
Matches 536; Conservative 0; Mismatches 288; Indels 15; Gaps 1;

QY 1586 TCTCTCACTTCAAGCATCTCCTACTCAGAAAATAGATAAAGAGAACTGTAACATCATG 1645
Db 1377 TGTCTCGGCACACAGCTGCTGCAGAAAATAAATCAAGGAGCCTGGGGAGGTTTC 1436

QY 1646 CTTCCGAGATGAGGATGACATCAATGATGTGACTTCTATGGCAGGGGTCAACCTTAATGA 1705
Db 1437 GTTTCGGGAGCATGATGACATTAATGATGTGATCGATGCTGAGTAAACTTGTCTAGA 1496

QY 1706 AGAAAATGCCGTGATCTTAGCAAACTCTGAATTTGTTGGCACACTCATTCAGTCAATG 1765
Db 1497 AGAAAGTGCAAGAAATATTAGCCAGCAACTCTGAAATTTGTTGGCACACTCATTCAG 1556

QY 1766 TAAAGATGAACCACTTCTTTTATTGGAGCTCTACAAAAGAGAACTCTTAGACATTTGT 1825
Db 1557 TAAAGATGAACCACTTCTTTTATTGGAGCTCTACAAAAGAGAACTCTTAGACATTTGT 1616

QY 1826 AAAGCATGACATTAACAGAACTTAACCTGTGATGCTGTGAACCTTGATCTCCCAAGCAACA 1885
Db 1617 AAACATGTTATACCGAAATTTACATCCAGATGTAGTAGTTATGTATCATCATGCCAC 1676

QY 1886 GGAACGACTAGGAGGCTTCTAGAAAACACTGCTGCAATTTGCTCAGCATTCGAATGACTAC 1945
Db 1677 ACAAGGCTACAGAACTCTGTAGAGAAATATCAGAAACAGCTCAGCAGAGAAACTTTTC 1736

QY 1946 TTCAAGGCAAGTCAAAATTTACATCTGTGTAGTATGATACCAAGTTCACAGCTCAAAATTTCT 2005
Db 1737 TTCAAGGATGACGACAGATATGAGCAGGCGAGTGTAGCTCCGGGCACAGCTCAAGTTT 1796

QY 2006 TGAAGAGTGGATCAATTTGGAGAGCAGAGAAAGGATTTTGAAGAAAGAGAAATCTTACT 2065
Db 1797 TGAACAGCTTGTCAATTCGAAAAGCAGAGGAGGATGAGCAGGCGGAGATCCTGTAT 1856

QY 2066 TAAGGCAAGCAAGAGTGTCTTAATAAGAGATCCAGACAGCTGAGATTTAAAGCAGAA 2125
Db 1857 GAGGCAAGCAAGGCTCTCGGTCAAGCAAGAGATCCAGAACTTAAGGCTGAAACAGAA 1916

QY 2126 AGCCAAAGAGTTACAGCAATTTGGAACCTTGACACATACAGATAGACAGCTTAATCTCAC 2185
Db 1917 GGCNAAGAGATGACGCAACAGGAACTTGGCAAAATAGACAGCGGGAGCCCAACCTCAC 1976

QY 2186 AGCTCTTGCAGCTATTGGCAAGCAAGAGAGAGACCACTAGA -----ATC 2230
Db 1977 AGCACTAGCAGCGATCGGGCCAGGAAAAGAGGAAAAGTGGACTGTCCGGGGCCGGGCTC 2036

QY 2231 TGGAAATTTAGGGCTTAAAGACAACTTCTTCTGTCTGGGACATCCAGCCTGACAGCCAC 2290
Db 2037 AGGAGCAGAGGGGTTCGGGCCCGGCTCAGTGTGTCCTCCAGGCACTCGGTGTCTGGAACCC 2096

QY 2291 CAAACAGTTGTCATCGTCAAGAAATCAGGAACTCTGCTCAGGACCTTGATATTGTTAT 2350
Db 2097 CAGCACTTTCAGGCAAGAGAAATCAGCGGGTCAACCTCAGGACCTCATATTGTTT 2156

QY 2351 GGAACAGGAACGGGAGATGAAGTATTCTCGAGCTCTATACCTGGCCCTTCTTGAAGTCA 2409
Db 2157 AGAAATGAACGTGAGACAAGCCATTTCATCTCTCTACAAAGCACTTCCCTTAAGTGAC 2215

| | | | |
|------------|--|--|-----------------|
| QY | 1766 | TAAAGATGAACCAATTTCTTTTATTGAGAGCTTACAAAAGAGAAATCTTAGACATTTGGTAA | 1825 |
| Db | 2595 | TAAAGATGAACCAATTTCTCTCCCAAGCGCTTTGCAGAGAAGAATATTAGAAATAGGTAA | 2654 |
| QY | 1826 | AAAGCATGACATTACAGAACTTAACCTGATCTGATGCTGTAAGTTCATCCCAAGCAACACA | 1885 |
| Db | 2655 | AAAACATGGTATACCGAATTTACATCCAGATGTAGTAAAGTATGTATCATCATGCCACGCA | 2714 |
| QY | 1886 | GGAGCTACGACGAGGCGCTTCTAGAAAACTGACTGCAATTTGCTCAGCATCGAATGACTAC | 1945 |
| Db | 2715 | ACAAAGCTACAGAACTTTGTAGAGAAATATACAGAAACAGCTCACAGAAAGAACTTTTC | 2774 |
| QY | 1946 | TTACAAGCAAGTCAAAATTACATCTGTGTAGTATACCAAGTTCACAGCTCAAAATTTCT | 2005 |
| Db | 2775 | TTACAAGGATGACGACAGATATGAGCAGGCGAGTACGCTCCGGGCACAGCTCAAGTTTTT | 2834 |
| QY | 2006 | TGAAAAGCTGGATCAATTTGGAGNAGCAGAAAGGATTTGGAAAGAAAGAAATGTTACT | 2065 |
| Db | 2835 | TGAACAGCTTGATCAATCGAAAAGCAGAGGAAGGATGACAGCGGAGATCCTGAT | 2894 |
| QY | 2066 | TAAAGCAGCAAGNGTCTTCTAATAAAGAAAGATCCAGAACAGCTGAGATTAAAGCAGAA | 2125 |
| Db | 2895 | GAGGCGAGCAAGTCTCGTCAAGACAGAAAGATCCAGAACAGTTAAGCTGNAACAGAA | 2954 |
| QY | 2126 | AGCCAAAGAGTTACAGCAATTTGGAATTTGCACAGATACAGCATAGACGCTAATCTCAC | 2185 |
| Db | 2955 | GGCAAGGAGATGCAGCAACAGGAACCTGGCACAATATGAGACAGCGGACGCCAACCTCAC | 3014 |
| QY | 2186 | AGCTCTTGAGCTATTGGACCAAGGAAAGAGACCACTAGA-----ATC | 2230 |
| Db | 3015 | AGCATTGAGCGGATCGGGCCAGGAAAGAGAAAGTGGACTGTCCGGGGCGCGGCTC | 3074 |
| QY | 2231 | TGGAATTTAGGGGCTTAAAGACAAACCTTCTTGCTTCTGGGACATCCAGCTGCAGACCCAC | 2290 |
| Db | 3075 | AGGAGCAGAGGGGTGGGCGCCCGCTCAGTGTCTCCAGGACGCTCGGGTGTCCGAACCCC | 3134 |
| QY | 2291 | CAAAAGTTGCATCGTCAAGAAATACAGAGAAATCTGCCCTCAGGAGCTTGATATTTTGTAT | 2350 |
| Db | 3135 | CAGACAGTTTCACGCGACAAAGAAATACAGGGGTTCACACCTCAGGAGCTCATATTTTGT | 3194 |
| QY | 2351 | GGAAAGGAACGGGAGATGAAGTATTCTGAGCTCTATACCTGGCCCTTCTGAAGTGA | 2408 |
| Db | 3195 | AGAAATGAACGTGAGACAAAGCAATTCAGTGTCTCTCAAGACATTCCTTAAAGTGA | 3252 |
| RESULT 11 | | | |
| AY038601 | | | |
| LOCUS | AY038601 | 2196 bp | linear |
| DEFINITION | Mus musculus TATA-binding protein associated factor TAFII135 mRNA, partial cds. | | ROD 29-AUG-2001 |
| ACCESSION | AY038601 | | |
| VERSION | AY038601.1 | | |
| KEYWORDS | | | |
| SOURCE | Mus musculus. | | |
| ORGANISM | Mus musculus | | |
| REFERENCE | | | |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| TITLE | Metsis M., Brunkhorst, A. and Neuman, T. | | |
| JOURNAL | Cell Type Specific Expression of the TFIID Component TAFII135 in the Nervous System | | |
| AUTHORS | Exp. Cell Res. (2001) In press | | |
| TITLE | 2 (bases 1 to 2196) | | |
| JOURNAL | Metsis M., Brunkhorst, A. and Neuman, T. | | |
| FEATURES | Submitted (06-JUN-2001) Medical Chemistry and Biochemistry, Karolinska Institute, Retzius vag 1, Stockholm 1717, Sweden | | |
| source | Location/Qualifiers | | |
| | 1..2196 | | |
| | /organism="Mus musculus" | | |
| | /strain="BALB/c" | | |
| | /db_xref="taxon:10090" | | |
| | /tissue_type="PO brain" | | |

| | | | |
|--|----------|---|---------------------|
| CDS | <1..1991 | | |
| /codon_start=3 | | | |
| /product="TATA-binding protein associated factor TAFII135" | | | |
| /protein_id="AAK94779.1" | | | |
| /db_xref="GI:15375062" | | | |
| /translation="ATLPTVLAPRLPQPKPNINQNPQLPPGMVLRSENGQLIMI | | | |
| VSQAALQMAHAQAQAPQSTMAPRPATPTGAPPVQISTVQAPGTPIIARQVPTTIIKQ | | | |
| VSQAALQMAHAQAQAPQSTMAPRPATPTGAPPVQISTVQAPGTPIIARQVPTTIIKQ | | | |
| AATETMENVKKCKSLSTLILKASSGKOSTETAAVNVKDLVNLQDKIEAFPTSRLY | | | |
| RELNSPQPYLYPFLKSLPALRLQTLTPDSNAITQSSQQQPPASQNTALTAVLSSS | | | |
| VORTAGKTAASVTSALQPPVLSLTQTVGVGKQAPPTPLVITQQPKPALRPPQVT | | | |
| LPTTPWALRQPHNRLMTLPOIQLNPVVPVPLPGLTALSTVSAQAAQAQK | | | |
| NKLKBPGGSGFRDDDDINDVASMGVNLSEESARILATNSLVLGTLTRCKDDTFLP | | | |
| APLQSTLEIKGKHGTTDCHPDVSVSHATQRLQNVVEKISETAQKNFSVKDDDR | | | |
| VQAASDVRQAQLKFFEQDQIEKQKDEQEREILLMRAKRSRQBPQEQRLQKTKEM | | | |
| QQQELAQHRQRDANLITAAIGFRKKRYDCTGPGSAGQSGFGAAVPGSGSGVTPRQ | | | |
| FTQRITRNLRLDLIFCLNENRERETSHSLLLYKAFIVTG* | | | |
| BASE COUNT | 597 a | 632 c | 564 g |
| ORIGIN | | | 403 t |
| Query Match | 12.7% | Score 325.8 | DB 10; Length 2196; |
| Best Local Similarity | 63.1% | Pred. No. 2.6e-74; | |
| Matches 526; Conservative | 0; | Mismatches 292; | Indels 15; Gaps 1; |
| QY | 1584 | ATTCTGTCTCACTTCAAGCATCTCTCTACTCAGAAAAATAGATAAAAGAGAAATGTAACATCA | 1643 |
| Db | 1146 | ACTGTATCTGCACAAGCAGCTGCTGCCAGAAAAACAACTCAAGGAGCGGGGGAGGC | 1205 |
| QY | 1644 | TGCTCCGAGATGAGATGACATCAATGATGTGACTTCTATGGCAGGGGTCAACCTTAAT | 1703 |
| Db | 1206 | TCCTTCGGGATGATGACATCAATGATGTTCGCTCAATGGCTGAGTGAACCTGTCT | 1265 |
| QY | 1704 | GAAGAAATGCCTGTCATCTAGCAACAACTCTGAATTTGGTTCGACACCTATTCAGTCA | 1763 |
| Db | 1266 | GAAGAAAGTGCAGAGTATTTGCCCAAACTCTGAGTTGGTGGGAGCTTGACCCGATCC | 1325 |
| QY | 1764 | TCTAAAGATGAACCAATTTCTTTTATTGGAGCTCTACAAAAGAGAAATCTTAGACATTTGT | 1823 |
| Db | 1326 | TGTAAGAGACACACCTTCTCTCTCCCGCGCTTTTCAGAGGAGACACACTGGAGATAGT | 1385 |
| QY | 1824 | AAAAAGCATGACATTTACAGAACTTAATCTGTGCTGTGACTGTGACTTCTCCCAAGCAACA | 1883 |
| Db | 1386 | AAAAAGCAGCGCATCAGGACTGCCATCCAGATGTAGTACGTTCAGTTACGTCTCATGCCACA | 1445 |
| QY | 1884 | CAGGAAGCACTACGAGCGCTTCTAGAAAACTGACTGCTCAATTTCTCAGCATCAATGACT | 1943 |
| Db | 1446 | CAGCAAGGCTGCAGAAATGTTGTAGAAAAATATCTGAAACGGCTCAGCAGAAAACTTC | 1505 |
| QY | 1944 | ACTTACAGCAAGTGAATAATTACATCTCTGTGTAGTATACCAAGTCCAGCTCAAAATTT | 2003 |
| Db | 1506 | TCTTATAAGGACGACGACAGATACGAGCAAGCGAGTGTCTCCGGCTCAGCTCAAAATTC | 1565 |
| QY | 2004 | CTTGAAGAGCTGGATCAATTTGGAGAGCAGAGAAAGGATTTGGAAGAAAGAGAAATGTTA | 2063 |
| Db | 1566 | TTTGAACAATGGATCAGATTTGAAAAGCAGAGAGAAAGATGAGCAGGAGGAGAGATCTTA | 1625 |
| QY | 2064 | CTTAAGCGACCAAGAGTCTGTTCTTAATAAGAGATCCAGAACGCTGAGATTAAAGCAG | 2123 |
| Db | 1626 | ATGCGGCGCAGGAGTCTCGATCCACAGAAAGAACCCAGAGCAATTAAGTTGAAGCAG | 1685 |
| QY | 2124 | AAAGCCAAAGAGTTACAGCAATTTGGAACCTTGCACAGATACAGCATAGAGACGCTAATCTC | 2183 |
| Db | 1686 | AAAAACAAAGAGATGACAGCAGCAGGAGCTGGCACAGATGCGACACAGATGCCAACCTC | 1745 |
| QY | 2184 | ACAGCTCTTGCAGCTATTGGACCAAGAGAGACAGACCA-----CTAGAA | 2228 |
| Db | 1746 | ACTGCACTGGCAGCATCGGGCCAGGAAAAAGAGAAAGTGGCACTGCACATGGGCCAGGA | 1805 |
| QY | 2229 | TCTGGAATTTAGGGCTTAAAGACAACTTCTTCTCTTCTGGGACATCCAGCTGCAGGCC | 2288 |
| Db | 1806 | TCAGGAGCAGGGCTCAGCGCCGGCGCAGCAGTCCCGAGCGGGTCCCGCGTGGGACCC | 1865 |
| QY | 2289 | ACCAACAGTTGATCGTCCAAAGAAATACAGAGAAATCTGCTCAGGAGCTTGTATTTTGT | 2348 |

Db 1866 CCCAGACAGTTACACGGCAAGATCACCGGTCACCTCAGGACCTCATATTTGT 1925
QY 2349 ATGGAACAGGAGCGGAGATGAGTATTCGAGCTCTATACCTGGCCCTTCT 2401
Db 1926 TTGAAGAACGAGCGTGAGCAAGCAATTCCTGTTGCTCTACAAAGCATTCCT 1978

RESULT 12
AP001197/c
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-775B10 map 18q11.2, WORKING
DRAFT SEQUENCE, 30 unordered pieces.
ACCESSION
AP001197
VERSION
AP001197.2 GI:8117643
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens DNA, clone:RP11-775B10.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 17422)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 174,222 genomic DNA of 18q11.2
Published Only in DataBase (2000)
2 (bases 1 to 17422)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (18-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7023969.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-775B10
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 151996 bases at least Q40
Consensus quality: 162600 bases at least Q30
Consensus quality: 167916 bases at least Q20
Insert size: 171322; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
30 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 16728 contig of 16728 bp in length
16829 31270 contig of 14442 bp in length
31371 45555 contig of 14185 bp in length
45656 59177 contig of 13522 bp in length
59278 70338 contig of 11061 bp in length
80210 89495 contig of 9772 bp in length
89596 96401 contig of 6806 bp in length
96502 102565 contig of 5906 bp in length
102566 109526 contig of 6861 bp in length
109527 109626 gap of 100 bp
115497 115596 contig of 5870 bp in length
115597 121502 contig of 5906 bp in length
121503 121602 gap of 100 bp
121603 126153 contig of 4551 bp in length
126154 126253 gap of 100 bp
126254 131007 contig of 4754 bp in length
131008 131107 gap of 100 bp
131108 135231 contig of 4124 bp in length
135232 135331 gap of 100 bp
135332 138844 contig of 3513 bp in length
138845 138944 gap of 100 bp
138945 142268 contig of 3324 bp in length
142269 142368 gap of 100 bp
142369 146076 contig of 3708 bp in length
146077 146176 gap of 100 bp
146177 149651 contig of 3475 bp in length
149652 149751 gap of 100 bp
149752 152991 contig of 3240 bp in length
152992 153091 gap of 100 bp
153092 157112 contig of 4021 bp in length
157113 157212 gap of 100 bp
157213 159965 contig of 2753 bp in length
159966 160065 gap of 100 bp
160066 161994 contig of 1929 bp in length
161995 162094 gap of 100 bp

| | | | | |
|---|--------|-------------------|---------|-----------|
| * | 162095 | 163533: contig of | 1439 bp | in length |
| * | 163334 | 163633: gap of | 100 bp | |
| * | 163634 | 165427: contig of | 1794 bp | in length |
| * | 165428 | 165527: gap of | 100 bp | |
| * | 165528 | 167111: contig of | 1584 bp | in length |
| * | 167112 | 167211: gap of | 100 bp | |
| * | 167212 | 169213: contig of | 2002 bp | in length |
| * | 169214 | 169313: gap of | 100 bp | |
| * | 169314 | 170970: contig of | 1657 bp | in length |
| * | 170971 | 171070: gap of | 100 bp | |
| * | 171071 | 173121: contig of | 2051 bp | in length |
| * | 173122 | 173221: gap of | 100 bp | |
| * | 173222 | 174222: contig of | 1001 bp | in length |

| Query Match | 11.58; | Score 294.6; | DB 2; | Length 174222; |
|-----------------------|---|--------------------|-----------|----------------|
| Best Local Similarity | 98.28; | Pred. No. 3.2e-66; | | |
| Matches 31; | Conservative 0; | Mismatches 4; | Indels 2; | Gaps 2; |
| QY 2236 | TTGAGGGCTTTAAAAGACAACTTCTTGGTGACATCCAGCTGACAGCCACCAAC | 2295 | | |
| Db 174026 | TTTAGGGCTTTAAAAGACAACTTCTTGGTGACATCCAGCTGACAGCCACCAAC | 173967 | | |

| | | | |
|------------|--------|--|--------|
| QY | 2296 | AGTTGCATCGTCCAAGATCAGAGAATCTGCCTCAGGAGCTTGATATTTTGTATGGAAC | 2355 |
| Db | 173966 | AGTTGCATCGTCCAAGATCAGAGAATCTGCCTCAGGAGCTTGATATTTTGTATGGAAC | 173907 |
| QY | 2356 | AGGAACGGAGATGAAGTATTTCTCAGCTCTATACCTGGCCCTTCTGAAGTGACCACTCC | 2415 |
| Db | 173906 | AGGAACGGAGATGAAGTATTTCTCAGCTCTATACCTGGCCCTTCTGAAGTGACCACTCC | 173847 |
| QY | 2416 | ACTCTTCCATCCATCTCTGCTATTACTGCGAAAGACACAAAAGCATTTGTTGCACT | 2475 |
| Db | 173846 | ACTCTTCCATCCAGATCTCTGCTATTACTGCGAAAGACACAAAAGCATTTGTTGCACT | 173787 |
| QY | 2476 | GTCTGAAATTTCAATTTCTGGAATAA-CACCAACATGAAGAGCATTGTTACGATT | 2534 |
| Db | 173786 | GTCTGAAATTTCAATTTCTGGAATAATCACCAACATGAAGAGCATTGTTACGATT | 173727 |
| QY | 2535 | AG-AACTTTATTAACCTTACTAT | 2558 |
| Db | 173726 | AGAACTTTATTAACCTTACTAT | 173702 |
| RESULT | 13 | | |
| AP001096 | | | |
| LOCUS | | 175553 bp DNA linear HTG 13-JUL-2000 | |
| DEFINITION | | Homo sapiens chromosome 18 clone RP11-813H9 map 18q11.2, WORKING DRAFT SEQUENCE, 12 unordered pieces. | |
| ACCESSION | | AP001096 | |
| VERSION | | AP001096.4 GI:9188475 | |
| KEYWORDS | | HTG; HTGS_PHASE1; HTGS_DRAFT. | |
| SOURCE | | Homo sapiens DNA, clone:RP11-813H9. | |
| ORGANISM | | Homo sapiens | |
| REFERENCE | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| AUTHORS | | 1 (bases 1 to 175553) | |
| TITLE | | Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y. | |
| JOURNAL | | Homo sapiens 175,553 genomic DNA of 18q11.2 | |
| REFERENCE | | Published Only in DataBase (2000) | |
| AUTHORS | | 2 (bases 1 to 175553) | |
| TITLE | | Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y. | |
| JOURNAL | | Direct Submission | |
| COMMENT | | Submitted (25-JAN-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, URL:htcp://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924) | |
| | | On Jul 14, 2000 this sequence version replaced gi:8117851. | |
| | | ----- Genome Center | |
| | | Center: RIKEN Genomic Sciences Center(GSC) | |
| | | Center code: RIKEN | |
| | | Web site: http://hgp.gsc.riken.go.jp/ | |
| | | Contact: hattori@gsc.riken.go.jp | |
| | | ----- Project Information | |
| | | Center project name: HumDraft18 | |
| | | Center clone name: RP11-813H9 | |
| | | ----- Summary Statistics | |
| | | Sequencing vector: PCR products; 100% of reads | |
| | | Chemistry: Dye-terminator ET-amersham; 100% of reads | |
| | | Assembly program: Phrap; version 0.990329 | |
| | | Consensus quality: 172065 bases at least Q40 | |
| | | Consensus quality: 173442 bases at least Q30 | |
| | | Consensus quality: 174017 bases at least Q20 | |
| | | Insert size: 17453; sum-of-contigs | |
| | | Quality coverage: 10.99x in Q20 bases; sum-of-contigs | |
| | | ----- | |
| | | NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and the order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved | |

```

1      65994 contig of 65994 bp in length
66095 80783 contig of 14675 bp in length
80870 95586 contig of 14717 bp in length
95687 108410 contig of 12724 bp in length
108511 122077 contig of 13567 bp in length
122178 135900 contig of 13723 bp in length
136001 149231 contig of 13231 bp in length
149332 158870 contig of 9539 bp in length
158971 165511 contig of 6541 bp in length
165612 171812 contig of 6201 bp in length
171913 173894 contig of 1982 bp in length
173985 175553 contig of 1559 bp in length.

* NOTE: this is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
* be preserved.

```

Faró, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meidrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (10-Apr-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 182884)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collumore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faró, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hags, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meidrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (22-Apr-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 10, 2002 this sequence version replaced gi:16041571.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L954
Center clone name: 9_E_17
----- Location/Qualifiers
1. 182884
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-9E17"
/clone_lib="RP11-9E17 Human Male BAC"
complement(711..1245)
/rpt_family="L2"
complement(2401..2496)
/rpt_family="MIR"
2522..2595
/rpt_family="L2"

repeat_region
complement(711..1245)
repeat_region
complement(2401..2496)
repeat_region
2522..2595
/rpt_family="L2"

FEATURES

source

repeat_region
complement(711..1245)
repeat_region
complement(2401..2496)
repeat_region
2522..2595
/rpt_family="L2"

repeat_region
3307..3341
/rpt_family="(TTC)n"
complement(4047..4535)
/rpt_family="LIME3A"
complement(4697..4951)
/rpt_family="AluSx"
complement(6008..6119)
/rpt_family="L2"
6391..6431
/rpt_family="MIR"
complement(6432..6737)
/rpt_family="AluSx"
6738..6916
/rpt_family="MIR"
complement(7019..7142)
/rpt_family="LIME10"
complement(7157..7262)
/rpt_family="L2"
9285..9308
/rpt_family="AT_rich"
9584..9620
/rpt_family="AT_rich"
9772..9897
/rpt_family="Charlie8"
9979..10250
/rpt_family="AluJo"
10254..10542
/rpt_family="AluSg"
10544..10568
/rpt_family="(TG)n"
complement(10758..11031)
/rpt_family="AluJo"
complement(11044..11392)
/rpt_family="AluSx"
12404..12555
/rpt_family="MIR"
complement(12537..12575)
/rpt_family="MIR"
complement(12608..12877)
/rpt_family="Tigger3(Golem)"
complement(12962..13263)
/rpt_family="AluSx"
complement(14519..14584)
/rpt_family="LIME4A"
complement(15397..15834)
/rpt_family="MLT1C"
complement(17303..17387)
/rpt_family="L3"
complement(17510..17606)
/rpt_family="MIR"
17613..17795
/rpt_family="MLT1J2"
complement(17796..18091)
/rpt_family="AluJb"
18092..18335
/rpt_family="MLT1J2"
complement(18345..18495)
/rpt_family="MIR"
complement(18522..18778)
/rpt_family="L3"
18861..18938
/rpt_family="MER5A"
complement(18939..19110)
/rpt_family="AluSg/x"
19112..19136
/rpt_family="(TG)n"
19133..19148
/note="Single clone coverage"
complement(19137..19182)
/rpt_family="LTR66"
19183..19283
/rpt_family="(TA)n"
complement(19284..19417)

repeat_region

```
repeat_region /rpt_family="AluJb"
complement(19418..19730)
repeat_region /rpt_family="AluSx"
complement(19731..19875)
repeat_region /rpt_family="AluJb"
19876..19954
repeat_region /rpt_family="MER5A"
complement(19973..20322)
repeat_region /rpt_family="L3"
20465..20888
repeat_region /rpt_family="MLT1B"
20986..21136
repeat_region /rpt_family="MER5A"
complement(21241..21337)
repeat_region /rpt_family="MIR3"
complement(21690..21817)

Query Match 11.5% Score 294.6; DB 9; Length 182884;
Best Local Similarity 98.2%; Pred. No. 3.2e-66;
Matches 319; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Qy 2236 TTGAGGGCTTAAAGACAACTTCTGCTTCTGGACATCCAGCTGACAGCCACCAAC 2295
Db 7466 TTTAGGGCTTAAAGACAACTTCTGCTTCTGGACATCCAGCTGACAGCCACCAAC 7525

Qy 2296 AGTTGCATCGTCCAGATACAGAGATCTGCCTCAGGACTTGATATTTTGTATGGAC 2355
Db 7526 AGTTGCATCGTCCAGATACAGAGATCTGCCTCAGGACTTGATATTTTGTATGGAC 7585

Qy 2356 AGGAACGGGAGATGAATGATCTCGAGCTCTATACCTGGCCCTCTCTGAAGTGACACTCC 2415
Db 7586 AGGAACGGGAGATGAATGATCTCGAGCTCTATACCTGGCCCTCTCTGAAGTGACACTCC 7645

Qy 2416 ACTCTTCATCCATCCTCTGCTATTTACTGCCAAGAACACAAAGCATTTGTGCACT 2475
Db 7646 ACTCTTCATCCATCCTCTGCTATTTACTGCCAAGAACACAAAGCATTTGTGCACT 7705

Qy 2476 GTCCTGAAATTCATTTCTGGAATAA-CACCAACATGAAGAGCATTTTACGATT 2534
Db 7706 GTCCTGAAATTCATTTCTGGAATAATCACCACATGAAGAGCATTTTACAGTT 7765

Qy 2535 AG-AACCTTTTAACTCTTACCTAT 2558
Db 7766 AGAACTTTTAACTCTTACCTAT 7790

RESULT 15
AP002752/c 190782 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 18 clone RP11-802C10 map 18q11.2, WORKING
DEFINITION DRAFT SEQUENCE, 38 unordered pieces.
ACCESSION AP002752
VERSION AP002752.1 GI:9188591
KEYWORDS HTG; HTGS, PHASE1; HTGS, DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-802C10.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 190782)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 190,782 genomic DNA of 18q11.2
JOURNAL Published Only in database (2000)
REFERENCE 2 (bases 1 to 190782)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(e-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
----- Genome Center
COMMENT
```

Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: <http://hgp.gsc.riken.go.jp/>
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-802C10
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 162369 bases at least Q40
Consensus quality: 175182 bases at least Q30
Consensus quality: 182192 bases at least Q20
Insert size: 187082; sum-of-contigs
Quality coverage: 4.15x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

| | | | | |
|--------|--------|-----------|-------|---------------|
| 1 | 14910 | contig of | 14910 | bp in length |
| 15011 | 26140 | contig of | 11130 | bp in length |
| 26241 | 36929 | contig of | 10689 | bp in length |
| 37030 | 48911 | contig of | 11882 | bp in length |
| 49012 | 57744 | contig of | 8733 | bp in length |
| 57845 | 68839 | contig of | 10995 | bp in length |
| 68940 | 77568 | contig of | 8629 | bp in length |
| 77659 | 87657 | contig of | 9989 | bp in length |
| 87758 | 93985 | contig of | 6228 | bp in length |
| 94086 | 100964 | contig of | 6879 | bp in length |
| 101065 | 107297 | contig of | 6233 | bp in length |
| 107398 | 113948 | contig of | 6551 | bp in length |
| 114049 | 120683 | contig of | 6635 | bp in length |
| 120784 | 126925 | contig of | 6142 | bp in length |
| 127026 | 130951 | contig of | 3926 | bp in length |
| 131052 | 134620 | contig of | 3569 | bp in length |
| 134721 | 138171 | contig of | 3451 | bp in length |
| 138272 | 143486 | contig of | 5215 | bp in length |
| 143587 | 148147 | contig of | 4561 | bp in length |
| 148248 | 152197 | contig of | 3950 | bp in length |
| 152298 | 155994 | contig of | 3697 | bp in length |
| 156095 | 159446 | contig of | 3352 | bp in length |
| 159547 | 162952 | contig of | 3406 | bp in length |
| 163053 | 165556 | contig of | 2504 | bp in length |
| 165657 | 168139 | contig of | 2483 | bp in length |
| 168240 | 170861 | contig of | 2622 | bp in length |
| 170962 | 173710 | contig of | 2749 | bp in length |
| 173811 | 175823 | contig of | 2013 | bp in length |
| 175924 | 177259 | contig of | 1336 | bp in length |
| 177360 | 178684 | contig of | 1325 | bp in length |
| 178785 | 180676 | contig of | 1892 | bp in length |
| 180777 | 182266 | contig of | 1450 | bp in length |
| 182367 | 184572 | contig of | 2206 | bp in length |
| 184673 | 186271 | contig of | 1599 | bp in length |
| 186372 | 188088 | contig of | 1717 | bp in length |
| 188189 | 189313 | contig of | 1125 | bp in length |
| 189414 | 190585 | contig of | 1172 | bp in length |
| 190686 | 190782 | contig of | 97 | bp in length. |

* NOTE: This is a 'working draft' sequence. It currently consists of 38 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 14910: contig of 14910 bp in length
* 14911 15010: gap of 100 bp
* 15011 26140: contig of 11130 bp in length

26141 26240: gap of 100 bp
 26241 36299: contig of 10689 bp in length
 36930 37029: gap of 100 bp
 37030 48911: contig of 11882 bp in length
 48912 49011: gap of 100 bp
 49012 57744: contig of 8733 bp in length
 57745 57844: gap of 100 bp
 57845 68839: contig of 10995 bp in length
 68840 68939: gap of 100 bp
 68940 77568: contig of 8629 bp in length
 77569 77668: gap of 100 bp
 77669 87757: contig of 9989 bp in length
 87758 87757: gap of 100 bp
 87758 93985: contig of 6228 bp in length
 93986 94085: gap of 100 bp
 94086 100964: contig of 6879 bp in length
 100965 101064: gap of 100 bp
 101065 107297: contig of 6233 bp in length
 107298 107397: gap of 100 bp
 107398 113348: contig of 6551 bp in length
 113949 114048: gap of 100 bp
 114049 120683: contig of 6635 bp in length
 120684 120783: gap of 100 bp
 120784 126925: contig of 6142 bp in length
 126926 127025: gap of 100 bp
 127026 130951: contig of 3926 bp in length
 130952 131051: gap of 100 bp
 131052 134620: contig of 3569 bp in length
 134621 134720: gap of 100 bp
 134721 138171: contig of 3451 bp in length
 138172 138271: gap of 100 bp
 138272 143486: contig of 5215 bp in length
 143487 143586: gap of 100 bp
 143587 148147: contig of 4561 bp in length
 148148 148247: gap of 100 bp
 148248 152197: contig of 3950 bp in length
 152198 152297: gap of 100 bp
 152298 155994: contig of 3697 bp in length
 155995 156094: gap of 100 bp
 156095 159446: contig of 3352 bp in length
 159447 159546: gap of 100 bp
 159547 162952: contig of 3406 bp in length
 162953 163052: gap of 100 bp
 163053 165356: contig of 2504 bp in length
 165357 165656: gap of 100 bp
 165657 168139: contig of 2483 bp in length
 168140 168239: gap of 100 bp
 168240 170861: contig of 2622 bp in length
 170862 170961: gap of 100 bp
 170962 173710: contig of 2749 bp in length
 173711 173810: gap of 100 bp
 173811 175823: contig of 2013 bp in length
 175824 175923: gap of 100 bp
 175924 177259: contig of 1336 bp in length
 177260 177359: gap of 100 bp
 177360 178684: contig of 1325 bp in length
 178685 178784: gap of 100 bp
 178785 180676: contig of 1892 bp in length
 180677 180776: gap of 100 bp
 180777 182266: contig of 1490 bp in length
 182267 182366: gap of 100 bp
 182367 184572: contig of 2206 bp in length
 184573 184672: gap of 100 bp
 184673 186271: contig of 1599 bp in length
 186272 186371: gap of 100 bp
 186372 188088: contig of 1717 bp in length
 188089 188188: gap of 100 bp
 188189 189313: contig of 1125 bp in length
 189314 189413: gap of 100 bp
 189414 190585: contig of 1172 bp in length
 190586 190685: gap of 100 bp
 190686 190782: contig of 97 bp in length.

Location/Qualifiers

source

1. .190782
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q11.2"
 /clone="RP11-802C10"

misc_feature

1. .14910
 /note="assembly_fragment"
 15011. .26140
 /note="assembly_fragment"
 26241. .36929
 /note="assembly_fragment"
 37030. .48911
 /note="assembly_fragment"
 49012. .57744
 /note="assembly_fragment"
 57845. .68839
 /note="assembly_fragment"
 68940. .77568
 /note="assembly_fragment"
 77669. .87657
 /note="assembly_fragment"
 87758. .93985
 /note="assembly_fragment clone_end:T7 vector_side:left"
 94086. .100964
 /note="assembly_fragment"
 101065. .107297
 /note="assembly_fragment"
 107398. .113948

Query Match

Best Local Similarity 11.5%; Score 294.6; DB 2; Length 190782;
 Matches 319; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 2236 TTGAGGGCTTAAAGACAACTTCTTGGCTCTGGGACATCCAGCCTGACAGCCACCAAC 2295
 || |||||
 Db 153393 TTGAGGGCTTAAAGACAACTTCTTGGCTCTGGGACATCCAGCCTGACAGCCACCAAC 153334
 QY 2296 AGTTGCATCGTCCAGAAATCAGGAAATCTGCTCAGGACTTGATATTTGTATGGAAC 2355
 |||||
 Db 153333 AGTTGCATCGTCCAGAAATCAGGAAATCTGCTCAGGACTTGATATTTGTATGGAAC 153274
 QY 2356 AGGACGGGAGATGAAGTATCTCGAGCTCTATACCTGGCCCTTCTGAAAGTGACCACTCC 2415
 |||||
 Db 153273 AGGACGGGAGATGAAGTATCTCGAGCTCTATACCTGGCCCTTCTGAAAGTGACCACTCC 153214
 QY 2416 ACTCTTCCATCCACATCCTTGGCTATTACTGCAAAAGACACAAAGCATTTGTCACCT 2475
 |||||
 Db 153213 ACTCTTCCATCCAGATCCTTGGCTATTACTGCAAAAGACACAAAGCATTTGTCACCT 153154
 QY 2476 GTCCGAAATTTCAATTTCTGGAATAA-CACCAACATGAAAGAGCATTTGTTAGGATT 2534
 |||||
 Db 153153 GTCCGAAATTTCAATTTCTGGAATAAATACCAACATGAAAGAGCATTTGTTAGGATT 153094
 QY 2535 AG-ACTTTTAACTCTTACCTAT 2558
 |||||
 Db 153093 AGAAACTTTTAACTCTTACCTAT 153069

Search completed: February 17, 2003, 02:40:41
 Job time : 9051 secs

FEATURES

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2003, 22:29:40 ; Search time 6475 Seconds
(without alignments)
11497.294 Million cell updates/sec

Title: US-09-763-909-1
Perfect score: 2558
Sequence: 1 gggaccctgtgacaaagt.....ctttattaaactttaacctat 2558

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_nam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result
No. | Score | Query
Match | Length | DB | ID | Description |
|---------------|--------|----------------|--------|----|------------|--------------------|
| 1 | 2534 | 99.1 | 2556 | 6 | I81227 | I81227 Sequence 1 |
| 2 | 2534 | 99.1 | 2556 | 9 | HSTAF11 | Y09321 H.sapiens T |
| 3 | 1145.2 | 44.8 | 182805 | 9 | AC017007 | AC017007 Homo sapi |
| 4 | 580 | 22.7 | 119081 | 2 | AC121320 | AC121320 Homo sapi |
| 5 | 580 | 22.7 | 169240 | 9 | AC022089 | AC022089 Homo sapi |
| 6 | 580 | 22.7 | 190782 | 2 | AP002752 | AP002752 Homo sapi |
| 7 | 338.2 | 13.2 | 3603 | 6 | I23466 | I23466 Sequence 15 |
| 8 | 338.2 | 13.2 | 3603 | 6 | I45759 | I45759 Sequence 15 |
| 9 | 338.2 | 13.2 | 4233 | 9 | HSU75308 | U75308 Human TBP-a |
| 10 | 337.2 | 13.2 | 3252 | 9 | HSTAF11.3 | Y11354 H.sapiens m |
| 11 | 325.8 | 12.7 | 2196 | 10 | AY038601 | AY038601 Mus muscu |
| 12 | 294.6 | 11.5 | 174222 | 2 | AP001197 | AP001197 Homo sapi |
| 13 | 294.6 | 11.5 | 175353 | 2 | AP001096 | AP001096 Homo sapi |
| 14 | 294.6 | 11.5 | 182884 | 9 | AC007996 | AC007996 Homo sapi |
| 15 | 294.6 | 11.5 | 190782 | 2 | AP002752 | AP002752 Homo sapi |
| 16 | 148.6 | 5.8 | 333300 | 2 | AC125091 | AC125091 Mus muscu |
| 17 | 148 | 5.8 | 16903 | 2 | AC016839 | AC016839 Homo sapi |
| 18 | 148 | 5.8 | 61882 | 2 | AC118057 | AC118057 Homo sapi |
| 19 | 147 | 5.7 | 277892 | 2 | AC125105 | AC125105 Mus muscu |
| 20 | 146.6 | 5.7 | 205466 | 2 | AC127767 | AC127767 Rattus no |
| 21 | 143.8 | 5.6 | 3535 | 3 | S63550 | S63550 transcripti |
| 22 | 142.2 | 5.6 | 3745 | 3 | AY069807 | AY069807 Drosophil |
| 23 | 142.2 | 5.6 | 4615 | 3 | DROTAFL10X | L06861 Drosophila |
| 24 | 142.2 | 5.6 | 4615 | 6 | I23459 | I23459 Sequence 1 |
| 25 | 142.2 | 5.6 | 4615 | 6 | I45752 | I45752 Sequence 1 |
| 26 | 137.6 | 5.4 | 205466 | 2 | AC127767 | AC127767 Rattus no |
| 27 | 124.4 | 4.9 | 62974 | 2 | AC118056 | AC118056 Homo sapi |
| 28 | 121.8 | 4.8 | 277892 | 2 | AC125105 | AC125105 Mus muscu |
| 29 | 112.4 | 4.4 | 158509 | 2 | AC114177 | AC114177 Rattus no |
| 30 | 109.8 | 4.3 | 62974 | 2 | AC118056 | AC118056 Homo sapi |
| 31 | 99 | 3.9 | 158509 | 2 | AC114177 | AC114177 Rattus no |
| 32 | 98.4 | 3.8 | 191918 | 2 | AC102692 | AC102692 Mus muscu |
| 33 | 94.6 | 3.7 | 121982 | 9 | HS1107C24 | AL109911 Human DNA |
| 34 | 92 | 3.6 | 333300 | 2 | AC125091 | AC125091 Mus muscu |
| 35 | 91.2 | 3.6 | 64035 | 2 | AC119227 | AC119227 Mus muscu |
| 36 | 89.6 | 2.5 | 194426 | 2 | AL663067 | AL663067 Mus muscu |
| 37 | 71 | 2.8 | 112277 | 2 | AC113239 | AC113239 Tetraodon |
| 38 | 62.4 | 2.4 | 2147 | 9 | AK097744 | AK097744 Homo sapi |
| 39 | 58.6 | 2.3 | 59865 | 2 | AC014422 | AC014422 Drosophil |
| 40 | 58.6 | 2.3 | 168479 | 3 | AC093454 | AC093454 Drosophil |
| 41 | 58.6 | 2.3 | 176056 | 3 | AC010066 | AC010066 Drosophil |
| 42 | 58.6 | 2.3 | 283821 | 3 | AE003528 | AE003528 Drosophil |
| 43 | 57.6 | 2.3 | 7218 | 6 | I66494 | I66494 Sequence 14 |
| 44 | 56.8 | 2.2 | 126712 | 9 | AL137077 | AL137077 Human DNA |
| 45 | 56.4 | 2.2 | 201470 | 2 | AC113649 | AC113649 Rattus no |

ALIGNMENTS

RESULT 1
I81227
LOCUS I81227 2556 bp DNA linear PAT 10-JUN-1998
DEFINITION Sequence 1 from patent US 5710025.
ACCESSION I81227
VERSION I81227.1 GI:3209517
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2556)
AUTHORS Dikstein, R. and Tjian, R.
TITLE Cell-type specific transcription factor
JOURNAL Patent: US 5710025-A 1 20-JAN-1998;
FEATURES Location/Qualifiers

This Page Blank (uspto)